J Euro Eco His.
ISSN : 0391-5115
Volume 26, Issue 1 (2026)
THE JOURNAL OF EUROPEAN ECONOMIC HISTORY

DEEPFAKES, DISINFORMATION, AND DEMOCRATIC DISCOURSE: REIMAGINING ARTICLE 19(1)(a) SAFEGUARDS IN THE DIGITAL AGE

Satyam Tomar
Assistant Professor,
TMCLLS,
Teerthanker Mahaveer University,
Moradabad.

ABSTRACT

The use of artificially generated deepfakes as a weapon poses a constitutional dilemma and finds no answers in the current Indian law. The dissemination of fake audiovisual material about public officials and all sorts of political actors, including candidates, not only damage reputations but also the discursive arena that is a prerequisite for democratic self-governance. While much has been written around the Puttaswamy privacy framework, the ShreyaSinghal proportionality matrix, and most recently the Digital Personal Data Protection Act, 2023, the response to how to calibrate free speech under Article 19(1)(a) of the Indian Constitution vis-a-vis its legitimate restrictions under Article 19(2) has not produced a coherent doctrinal framework. You are based everywhere until October 2023, and this article helps fill a major void in the present legal scholarship and legislative design.

This paper formulates three interrelated arguments using an interdisciplinary synthesis of democratic theory and communication sciences, a critical comparative approach as well as doctrinal constitutional analysis. To start with, deepfakesare a different form of speech, a sort of epistemic sabotage that very little fits into the normal definitions for incitement, sedition or defamation. Second, the current landscape of regulation represented by the Representation of the People Act 1951, DPDP Act 2023 and Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2021 is asymmetrical in nature and procedurally incoherent which fails to protect democracy's epistemic conditions for participatory engagement as put forth.

Third, the article presents the Threshold-Reasonableness-Intent-Discourse (TRID) Framework, which is a new doctrinal standard with four operationalized criteria to help legislators and courts determine whether any particular deepfake regulation passes constitutional muster without endangering the pluralism that Article 19(1)(a) seeks to protect.

The paper ends with a set of concrete institutional responses that provide legislators (or someone) with a roadmap toward conformity constitutional or near as practical impact, such as an amendment to Schedule III of the Representation of the People Act; and Deepfake Electoral Integrity Protocols.

Keywords: Deepfakes; Disinformation; Democratic Discourse; TRID Framework; Digital Personal Data Protection Act; AI Regulation; Electoral Integrity; Freedom of Speech; Synthetic Media.

J Euro Eco His.
ISSN : 0391-5115
Volume 26, Issue 1 (2026)
THE JOURNAL OF EUROPEAN ECONOMIC HISTORY

I. Introduction: Democracy in the Age of Epistemic Sabotage

A forged video clip of a prominent Indian opposition leader saying his party would be pulling out all its candidates from the general election went viral on WhatsApp and Telegram in March 2024, racking up more than four million views in just three days before fact-checkers stepped in. While the video was identified as an AI-generated deepfake, the average viewer could point out no errors in it. No one was prosecuted. No specific law was applicable. The incident was awkwardly illustrative of a deep contradiction in Indian constitutional democracy in the digital age, namely, that the structure of protections for free expression was built for a society where the citizens create, share and consume communicative acts. The machine learning algorithms on which the internet economy relies were not built for a world where robots could churn out what seemed like real political debate at scale, and disseminate it through ecosystems that optimize for emotionally engaging content rather than epistemically rigorous discourse.

The central tenet of this Article is that deepfake-mediated disinformation constitutes a mode of epistemic sabotage theory of harm both constitutionally novel and unmoored, as those familiar with the hundreds of judicial precedents since the founding will know, from any particular category within which Article 19 (1) encompasses speech perceived as harmful. Deepfakes are not your traditional misinformation where a false claim is made only. They produce evidence of the words and events, sense in which citizen use to build political beliefs. Deepfakes violently bend the sensory register of democratic happenings, whereas conventional defamation uses a lie to fundamentally gag an individual. Actually, where deepfakes exploit fictition so as to diminish independent political actors, incitation turns to words as a springboard for action.

The constitutional aspect of this challenge is not very clear. My legal education began in India and I have watched over the years as Indian courts developed a massive body of law on freedom of expression starting from Romesh Thappar v. State of Madras (1950) to Shreya Singhal v. Union of India (2015), and Anuradha Bhasin v. Union of India (2020) but no court ever adopted these artificial images or used synthetic media.

Intermediary Guidelines, 2021 imposes procedurally cumbersome and substantively weak traceability and takedown obligations; the DPDP Act 2023 addresses personal data but not democratic backsliding; while Section 66D of the IT Act,2000 inadequately defines impersonation. The attitudinal regulatory gap is such dramatically trimmed that the epistemic armour leaves in its wake is so broad, this paper shows, as to threaten the epistemic prerequisites of democratic self-governance itself.

This piece makes the case for a rebuilt constitutional base over an entirely new regulatory structure. The core of the argument is that, when read through the lens of transformative

constitutional theory and democratic constitutionalism, Article 19(1)(a) possesses at least some normative resources to respond to deepfake harms provided those resources are given effect through a coherent doctrinal standard capable of being operationalized by legislators and courts alike. The original contribution of this article is that standard (the Threshold-Reasonableness-Intent-Discourse (TRID) Framework), which was proposed in Section VII.

The rest of the paper is organised as follows. Section II describes the methodological framework and doctrinal-comparative-interdisciplinary approach used here. By mapping the current scholarly debate, Section III delineates the particular need that this essay is intended to fill. Section IV provides a technical and political-theoretical analysis of the harm caused by deepfakes. Section V analyzes the status quo of the Article 19(1)(a) infrastructure and its democratic values; Section VI examines India's [proposed] regulatory landscape in addition to a few other nations that find themselves in a similar dilemma. Section VII contains the TRID Framework that is proposed and developed. Section VIII addresses the institutional architecture for implementing the Framework. End of Section IX gives specific legislative recommendations.

II. Methodology: Doctrinal Analysis, Comparative Method, and Interdisciplinary Synthesis

To adequately address this topic, which lays at the intersection of constitutional law, information technology, and democratic theory the paper employs tripartite methodology. Such a methodology is appropriate for an issue that involves questions of normative theory (what should Article 19(1)(a) allow?), Indeed, what legislation is already on the books (what does it allow?), and institutional design (what are justifiable constraints?).

The predominant method is doctrinal constitutional analysis. Contemplating the relevant case law, Article 19's language, form and precedential elaboration as well as the constitutional standard employed by courts - proportionality, non-arbitrariness and democratic need. This is not merely descriptive exegesis; as Krishnaswamy (2009) describes best the Indian constitutional process: it is "interpretive construction," which relies on the moral commitments that animate the constitutional text and its internally coherent logic.

The second methodology is critical comparative analysis. This article systematically explores four comparative jurisdictions and they are European Union (EU AI Act and Digital Services Act), United Kingdom (S 15 Online Safety Act 2023), United States (First Amendment jurisprudence and the Defend Elections from Covert Harmful AI Activities Act) and Australia (Electoral Integrity legislation). The objective is not to facilitate transplantation, but to distill regulatory design options that can be constitutionally attuned with respect to the Indian context. The comparative methodology stems from Michaels' (2006) approach of "functional equivalency", which asks whether foreign regulatory structures fulfill functionally equivalent purposes in legally comparable situations.

The third is that of multidisciplinary synthesis. Where applicable, the constitutional analysis is informed by empirical evidence from computational media studies (about the technical trajectory of synthetic media generation), political science (about the quantifiable harms to

electoral integrity), and communication science (about the cognitive effects of deepfakes on information processing). This interdisciplinary interaction is not cosmetic; it will require operationalising the unfairness criterion in the TRID Framework with empirical specificity, rather than intuitive approximation to a "harm threshold" form.

III. Mapping the Existing Discourse: The Literature and Its Limitations

The academic literature on deepfakes, disinformation, and law has developed substantially since the publication of Chesney and Citron (2019) article, a California Law Review essay whose intervention remains the most cited in the field. Chesney and Citron constructed a responsibility framework drawn from US tort law and then systematically enumerated deepfakes' privacy, national security, and democratic vulnerabilities. Their framework, despite its analytical sophistication, is distinctly American in construction: it functions entirely within the parameters set by the essentially US First Amendment–based content-neutrality presumptions (which are much stronger than Indian constitutional law predicates); it rests on the functional capacities of private civil litigation; and ultimately it seeks to understand Section 230 immunity as a limiting design constraint for their model.

Citron and Chesney's (2019) Foreign Affairs essay, "Deep Fakes and the New Disinformation War," introduced the deepfake problem into the theatre of geopolitically charged information warfare, reaching a broader audience of policy-makers with their national security framing that is neglected by more domestic-focused legal analyses. However, its recommendations remain at the level of policy advocacy as opposed to doctrinal specification.

In consequence, Indian constitutional scholarship is evidently inattentive to the field. The best explanation of Article 19(1)(a) doctrine can be found in Bhatia's (2016) Offend, Shock, or Disturb, which demonstrates how the courts slowly began to formulate a proportionality-based approach to restrictions on speech that, though not formally embracing Alexy's theory of balancing conflict considerations, operates essentially as a balancing test. However, Bhatia's analysis does not address synthetic media and was written before deepfakes became widespread. This essay builds on, and elaborates Krishnaswamy 2009's Democracy and Constitutionalism in India which provides a theoretical framework for interpreting essay 19 in terms of democracy as an inherent element of its identity. But this does not cover AI or information warfare. Analysis of the impact of synthetic media on freedom of expression has been scarce in post-Puttaswamy scholarship (Narrain, 2018; Chander& Le, 2020), and found almost solely within the context of right to privacy and data protection.

For definitions of the regulatory environment, Balkin's (2018) "Free Speech in the Algorithmic Society" provides a theoretically sophisticated account of how algorithmic platforms have altered conditions for democratic speech. It also makes a compelling argument that First Amendment doctrine requires infrastructure regulation instead of merely speech regulation. There have been no systematic attempts to use his idea of “information fiduciary" in Indian constitutional contexts, although it has generated significant academic interest. While the Council of Europe’s Information Disorder framework (Wardle &Derakhshan, 2017) has so far

not been integrated within Indian constitutional analysis, it provides a typology widely used in European regulatory discourse that distinguishes between misinformation, disinformation and malinformation on axes of falsity and intention to harm.

Both Model Code of Conduct guidance(2019) from the Election Commission of India that prohibits "misleading" electoral material without addressing the evidentiary problems around identifying deepfakes, or countering these technologies' arms race between generation versus detection and a Data & Society report by Paris & Donovan (2019),"Deepfakes and Cheap Fakes" drawn attention to this specific electoral dimension of deepfakes.

Not a single studies in Indian constitutional law, comparative information law or technology policy has proposed either a structured doctrinal standard courts could apply under Article 19 that treats deepfakes as a unique kind of constitutional animal (as opposed to historical disinformation), or one which provides precise enough criteria so that case outcomes are predictable while regulatory overreach is avoided. This is the important gap that this article recognizes and seeks to address. The literature either applies frameworks from other jurisdictions without engaging with Indian constitutional doctrine (Balkin, 2018), describes the problem yet does not set out any constitutional standard (Chesney & Citron, 2019; Paris & Donovan, 2019), or deals with adjacent but distinct constitutional concerns (e.g., privacy Narrain, 2018; surveillance Rao, 2021; platform liability Sinha, 2020).The TRID Framework proposed in Section VII is designed to fill this gap.

IV. The Anatomy of Epistemic Sabotage: Technical Dimensions and Democratic Consequences

A. The Technical Architecture of Deepfakes

According to Tolosana et al. (2020), deepfakes are artificial media artefacts created by deep learning techniques, usually Generative Adversarial Networks (GANs) or diffusion models that generate audio-visual footage showing people acting out or saying things that never happened. Although imprecise, the term captures a technologically continuous phenomenon: at the high-fidelity end, state-of-the-art GAN architectures can produce synthetic video of sufficient quality to fool trained forensic analysts in controlled laboratory conditions; at the low-fidelity end, "cheap fakes" involve simple editing, speed alteration, or decontextualisation of authentic media (Farid, 2022).

There are three components of the technology of deepfakes that link directly to the constitution. One, accessibility: the computing resources needed to generate realistic deepfakes are now available to anyone who can acquire reasonably high-end consumer hardware along with freely-available open-source software formerly limited state actors and well-resourced adversaries. Production costs, which once constrained the scale of advanced information operations, have effectively vanished (Chesney & Citron 2019:1764). Secondly, scale: deepfake production might be automated, potentially allowing for the mass generation of fake content against multiple targets at once. Finally, the asymmetrical detectability of deepfakes (Farid, 2022) creates a structural epistemic disadvantage for defenders as contrasted

with attackers. The legal framework must account for this asymmetry; a regulatory regime premised on rapid detection and removal will be perpetually inadequate.

B. Democratic Harms: Beyond Defamation

The current doctrinal categories of incitement, sedition, and defamation are insufficient to fully reflect the constitutional relevance of deepfakes. These categories deal with threats to one's physical safety, political order, and reputation, respectively. The contamination of the epistemic environment that democratic discourse takes place is a completely separate harm caused by deepfakes.

A helpful theoretical framework is provided by Habermas's (1989) description of the public sphere, which has been revised for digital circumstances. He claims the existence of communication conditions that enable citizens to come up, revise and articulate their political opinions through reasonable debate is the backbone of democratic legitimacy. For these conditions to be reflected in the lives of citizens we must, at the very least, be able to tell apart real political communication from fake. Deepfakes systematically weaken this ability. The cognitive effects of exposure to misleading information endure even after explicit corrections, as Lewandowsky et al. (2017) show through experimental psychology study; they refer to these phenomena as "the illusory truth effect".

Deepfakes exploit and magnify this effect by tying false beliefs to what is represented as direct perceptual experience the evidentiary force of actually "seeing" or "hearing," rather than propositional claims that can be subjected to traditional standards of epistemic scrutiny.
Sunstein (2017) adds some evidence for this explanation in the form of his analysis of the "echo chamber" effect. This means that deepfakes disseminate preferentially over partisan media networks operating in algorithmically curated information environments, and where fact-checking corrections cannot penetrate. This culminates in what Rini calls "partisan epistemology," or the systematic adjustment of epistemic norms to political party. As such, the democratic injury is not merely misinformation but the erosion of the shared epistemic norms that ground political conflict, rather than information warfare.

In the electoral context, deepfakes lead to three classes of compounded harm (Election Commission of India, 2019). First, they enable candidates to create comments that alter how voters feel on the basis of imaginary events. Second, they exploit careful ambiguity that renders even real statements of political import contestable as "deepfakes," undermining the potential power of accountability systems. Ultimately, this so-called "liar's dividend" (Chesney & Citron 2019: 1796) may be just as harmful as the fabrication. Third, deepfakes act as another of what Deibert (2019) characterises as the "chokepoint" processes by which authoritarian government is produced in nominally democratic settings through eroding institutional confidence in electoral systems generally and against specific candidates or parties.

J Euro Eco His.
ISSN : 0391-5115
Volume 26, Issue 1 (2026)
THE JOURNAL OF EUROPEAN ECONOMIC HISTORY

V. The Article 19(1)(a) Architecture: Constitutional Freedoms and Their Democratic Premise

A. The Democratic Foundation of Free Expression

Article 19(1)(a) of the Indian Constitution guarantees every citizen of India the right to freedom of speech and expression. Through seven decades of constitutional jurisprudence, the Supreme Court of India has turned this promise into a constitutionally guaranteed fundamental right not merely as an administrative protection against government censors but also as a substantively democratic fundamental obligation. From RomeshThappar v. State of Madras (AIR 1950 SC 124), which established the basic presumption against any state restriction, to Sakal Papers (P) Ltd. v. Union of India (AIR 1962 SC 305), which recognised the economic dimensions of press freedom; Bennett Coleman & Co. v. Union of India ((1973) 2 SCC 788), in which the Court unequivocally identified press freedom with democratic pluralism and ultimately to ShreyaSinghal v. Union of India ((2015) 5 SCC 1), where the constitutional requirement of proportionality was operationalised with analytical precision.

The democratic assumption behind Article 19(1)(a) is not coincidental. As Krishnaswamy (2009, p.67) demonstrates through close textual and historical analysis, those who framed the document that became the draft Constitution viewed freedom of expression as constitutively necessary to democratic self-governance: Citizens must speak, criticize, and deliberate if mechanisms of representative democracy are to function as instruments of popular sovereignty rather than instruments of elite management. This basic democratic principle is what makes the deepfake problem such a big deal. Assuming that proper citizen access to correct political information and deliberative speech are preconditions of democratic self-governance, we find a partial protection for speech in Article 19(1)(a). Then speech that systematically destroys the accuracy of political information and corrupts deliberative conditions stands in a different constitutional position from speech that merely offends, disturbs, or challenges.

B. The Proportionality Framework and Its Limits

The proportionality framework articulated in Shreya Singhal offers a procedural structure for evaluating what goal deepfake regulation serves one whose speech restrictions found under Article 19(2) must not only have a legitimate aim but also must naturally correlate with that aim and be the least restrictive means available. This paper explains that the ShreyaSinghal framework is necessary but not sufficient for deepfakes. The requirement for the proportionality inquiry is that this communicative material to be restricted is recognised and shown to exist by the court. Specifically, this assumption is rendered more complicated in three dimensions with deepfakes.

First, the communicative content of a deepfake is always deceptive, treating what is generated expression as if it were genuine. To assess the appropriateness of such a restriction on this content, it must be possible to ascribe communicative content to its apparent speaker; but when that speaker did not in fact produce it, do we have access necessary for such an epistemic

attribution. Second, Now the impact of deepfakes is systemic, not personal. Since the proportionality analysis was largely designed with individual expression in mind, it has not been adapted to capture collective epistemic harms of large synthetic media campaigns. Third, the specific type of harm that this article identifies as the equivalent of "epistemic sabotage" is separately defined from any of the conventional Article 19(2) grounds: sovereignty, integrity, security, public order, decency and morality, contempt of court / defamation and incitement.
As a result, the distinction between the proportionality framework and the issue of deepfakes is doctrinal, rather than purely empirical. It requires the establishment of a constitutional standard that clearly recognizes the type of speech harm that deepfakes produce; it cannot be cameared in improved fact-finding in discrete cases.

C. The Right to Receive Information and the Epistemic Dimension of Article 19

An integral part of the Article 19(1)(a) doctrine which is relevant to deepfake issues, is the right to receive information. This right was recognised by the Supreme Court in Indian Express Newspapers (Bombay) Pvt. Ltd. v. Union of India AIR 1986 SC 515, and build upon in Secretary, Ministry of Information and Broadcasting v Cricket Association of Bengal (1995) 2 SCC 161. The right to information is the counterpart of the right to expression which guarantees the audience's interest in genuinely varied communicative content. While this correlative right is largely rooted in monopolistic media control and governmental censorship, it seems to intuitively translate well into the deepfake context.

In fact, deepfakes are not attempts to suppress information availability but rather to corrupt its integrity: instead of limiting the supply of information they increase it. So, when a citizen is simply 'receiving' fake audio-visual content from a political actor, this does not amount to information in the sense of a constitutionally-protected activity, but misinformation that systematically distorts rather than assists on the formation of political preferences. In particular, if the right to receive information is or can be deemed a safeguard on the epistemic conditions of democratic participation, it follows that Article 19(1)(a) does not constitutionally protect a systematic degradation of those very requirements by technologies conditioning widespread false perceptual experience. Rather, it is a medium of expression that the state has the power to regulate--and arguably even an obligation to regulate under its constitution.

VI. The Regulatory Lacuna: India's Inadequate Legal Response

A. The Information Technology Act, 2000 and Its Amendments

Section 66D of the Information Technology Act, 2000, makes it an offence to cheat by personation using a computer resource (maximum punishment:3 years of imprisonment with fine). On the surface, this clause would apply to a deepfake of a public person. But Section 66D is not the end of the road, because it fails on three theological counts. First, it includes a definition of "cheating" in the provision that refers to Section 415 of the Indian Penal Code (IPC), including the requirement that the deception should lead to "wrongful gain or wrongful loss to the person deceived or to any other person." Because individual financial injury may be impossible to show from a politically-motivated deepfake, the electoral harm is left without

any prosecutorial hook.Second, Section 66D has not yet been authoritatively interpreted to include AI-generated synthetic media as it was written specifically aimed at preventing human impersonation for commercial gain. Third, even if it does apply there are two problems, Section 66D represents a post-hoc individual remedy rather than the systemic-solution-prevention-reaction that we need in order to address the platform-level systemic damages that make deepfakes democratically risky.

The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (the Intermediary Guidelines) set out various obligations on "Significant Social Media Intermediaries", including traceability of message originators and prompt removal of content that is notified to be illegal. But as Sinha (2020) illustrates, the traceability requirement was originally designed to combat end-to-end encryption; it poses both analytical and technical contradictions with the deepfake problem: the central challenge of deepfakes is not identifying who disseminates a video –but forensically establishing its synthetic provenance a challenge that finds no redress under the Intermediary Guidelines. Also, while the 36-hour to remove limit is quick by more traditional standards, when it comes to viral deepfake propagation which can accrue millions of views within hours of being seeded even then it's not enough.

B. The Digital Personal Data Protection Act, 2023: A Partial Remedy

Arguably, the most significant recent legislative initiative in the sphere of Indian digital governance is embodied in the Digital Personal Data Protection Act, 2023 (DPDP Act). It is sparse, but it carries genuine relevance to the deepfake problem. The Act forbids the processing of personal data unless there is a consent and most deepfakes require unconsented use of the subjects biometric identifiers (their voice, facial features, etc. The Act's consent framework could potentially provide a private right of action for subjects of non-consensual deepfakes.

However, the DPDP Act's remedial architecture, centred on the Data Protection Board and administrative penalties, is not designed to address the democratic harms of electoral deepfakes. The subject of a deepfake electoral video may have a private data protection remedy; the democratic public, whose deliberative capacity has been distorted, has none. As Zuboff (2019) observes in the analogous context of surveillance capitalism, "the harm is not merely done to individuals but to the architecture of democratic life" (p. 201), and architecturally democratic harms require architecturally democratic remedies.

C. Electoral Law and the Representation of the People Act, 1951

Section 123(4) of the Representation of the People Act, 1951 designates as a "corrupt practice" the publication of statements that are false and that the candidate knows or believes to be false, regarding the personal character or conduct of any candidate, when made for electoral purposes. It is a strict but imperfect restriction-is the most similar enactment of an election law to deepfake regulation presently in effect. There is no strict liability that could be used to deter an automated deepfake distribution campaign; a publisher is not liable unless they

"know or believe" the statement is false. This clause only protects remarks about a person's own "character or conduct"; fake election announcements, the pronouncement of policies and other politically damaging synthetic speech is omitted. The lack of criminal penalties means it offers inadequate deterrent value, as civil electoral consequences can be balanced against possible sanctions by the IEB. It also empowers the Election Commission of India, which does not have fraud forensic expertise needed to investigate the large-scale legal validity of deepfakes.

D. Comparative Perspectives: Regulatory Design Choices

Insights on this design choice come from a comparison of deepfake-specific regimes across four jurisdictions. The EU's proposed AI Act (2024) has a risk-based approach, treating political and electoral deepfakes as "high risk" AI applications requiring ex ante conformity assessment and third-party auditing, and putting "deep fakes" in the entertainment domain into the classification of systems subject to mandatory disclosure obligations (EU AI Act, Article 50(4)). The Digital Services Act of 2022 (DSA) follows up by requiring Very Large Online Platforms to conduct systemic risk assessments related to "negative effects for civic discourse" and take action based on those findings (DSA, Article 34). Together, these tools create an upstream set of rules targeting the enabling infrastructure for deepfake production and distribution rather than individual instances.

Meanwhile the UK Online Safety Act 2023 uses an outright different approach and titled "primary priority harmful content" in a broad classification of such as "disinformation" without any deepfake limitation whatsoever. Under their existing statutory "safety duties", platforms must conduct risk assessments and implement appropriate safety measures. The structural innovation of the Act to impose systemic obligations on platforms rather than ad-hoc decisions on specific pieces of content provides a model that is relevant in the Indian context, even if critiques have emerged about its heavy reliance on self-regulation by platforms and absence of covering electoral deepfakes explicitly.

The categorical nature of U.S. content regulation per the First Amendment makes comprehensive deepfake regulation a challenge. California state law that prohibits deepfakes within 60 days of elections, AB 730 (2019) was the first of its kind but successfully challenged as unconstitutional. This U.S. approach, due to its First Amendment constraints, provides a cautionary tale of the dangers of lacking sufficient upstream regulation: downstream remedies are necessarily more intrusive and constitutionally tenuous than an appropriately designed system mandate (Balkin 2018, 1187).

The comparative analysis supports three design principles for an Indian regulatory framework: 1) systemic upstream regulation of platforms and production tools rather than mere reactive content-by-content adjudication; 2) mandatory disclosure obligations tied to technical provenance standards (Content Authenticity Initiative metadata); and 3) a robust enforcement mechanism with adequate forensic skills. The TRID Framework suggested below puts these ideas into practice within the limits of Indian constitutional doctrine.

J Euro Eco His.
ISSN : 0391-5115
Volume 26, Issue 1 (2026)
THE JOURNAL OF EUROPEAN ECONOMIC HISTORY

VII. The TRID Framework: A Reconstituted Constitutional Standard for Deepfake Regulation

The most significant analytical contribution of the article is the formulation of a four-pronged doctrinal standard, called the Threshold-Reasonableness-Intent-Discourse (TRID) Framework, for analysing governmental actions that limit or regulate deepfake expression under Articles 19(1)(a) and 19(2). The Framework is designed for optional operationalization by legislators drafting regulatory tools, as well as courts determining the constitutionality of laws governing deepfakes. It is, without doubt, rooted in the conceptual paradigm of Indian constitutional law but extends and strengthens its foundations to also address the distinctive dimensions of prejudice wrought by synthetic media.

Prong 1: Threshold of Harm (T) - The Epistemic Nexus Test

The first component requires that a regulatory action regulate communicative content that meets the markup bar of epistemic harm: it must be capable of meaningfully hindering citizens' capacity to make informed political decisions concerning candidates, elections or public institutions. While there may be arguments in favor of limiting other forms of merely offensive, embarrassing or false speech that do not interfere with democratic deliberation, deepfake regulation is instead geared towards undermining the epistemic conditions for democratic participation; this "epistemic nexus," is what distinguishes such limits from those on more mundane offensiveness.

The epistemic nexus test has three components. First, subject-matter: the content must concern electoral processes, candidates, political parties, or the actions of public institutions. Deepfakes targeting private individuals without public relevance, commercial fraud, or non-political social harms fall outside this constitutional category, though they may be addressed by other regulatory regimes. Second, falsity: the content must be synthetic in a verifiable manner; it should depict or impersonate statements, appearances, or events that did not literally occur. This aspect underscores the vital requirement for institutional capabilities in the enforcement framework and urges courts to engage with technical forensic information. Third, reach: the content must have achieved, or been obviously intended to achieve, sufficient circulation that it might potentially affect political discourse or outcomes across the electorate. This "reach" condition prevents the Framework from being used as a blunt instrument against truly marginal or non-political synthetic content. They are fine-tuned to the particular kind of democratic harm that election-related deepfakes can inflict.

The epistemic nexus test finds doctrinal support in the Supreme Court's recognition, in AnuradhaBhasin v. Union of India ((2020) 3 SCC 637), that restrictions on expression must be "proportionate to the harm sought to be prevented" and that the harm must be "direct and proximate" rather than speculative. This article proposes that "direct and proximate harm to democratic deliberation" constitutes a legitimate ground for restriction under Article 19(2)'s

"public order" and "sovereignty and integrity" heads, broadly construed. The constitutional basis is amplified by the Puttaswamy Court's recognition that privacy "enables individual self-determination," which, in the democratic context, encompasses the epistemic self-determination of voters (Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1, paragraph 169).

Prong 2: Reasonable Attribution Standard (R) — The Provenance Precision Requirement

The second prong addresses the attribution challenge is the structural epistemic difficulty of determining whether content is synthetic and requires that regulatory enforcement deploy technically reasonable and procedurally auditable methods of establishing the synthetic origin of contested content. However, this constitutional requirement equipping the state with authority to act pursuant to contentious or inadequate forensic evidence is open to abuse by totalitarian regimes: the legitimate political expression in society could be promulgated as "deepfake disinformation" by the state why a measure of genuine dissent will be suppressed.
The Reasonable Attribution Standard requires three procedural safeguards. First, the regulatory body needs to have or employ technically qualified forensic expertise, deploying techniques that conform to current best-practice standards such as those developed by the IEEE Media Provenance Working Group, the Content Authenticity Initiative and equivalent international technical bodies. Second, attribution findings must be subject to independent review through a non-executive body, this article proposes a Deepfake Adjudication Panel, addressed in Section VIII before regulatory action is taken, except in cases of genuine imminence risk to electoral integrity where emergency interim measures are warranted. Third, the target of a regulatory attribution finding must have effective notice and an opportunity to contest the finding through an expedited procedural mechanism. This third safeguard reflects the Supreme Court's insistence in ShreyaSinghal((2015) 5 SCC 1, paragraph 107) that procedural safeguards are constitutionally integral to restrictions on expression.

The Supreme Court, of course, held that the restrictions on expression must be backed by "public safety or public order" to be valid (S. Rangarajan v P Jagjivan Ram (1989) 2 SCC 574, para 30) and they cannot be "like a match to a powder keg"; this is somewhat analogous with what I have called the Reasonable Attribution Standard above. Proof on a structural level, as opposed to only at the formal level is therefore required by constitutional limitation. In deepfakes, this implies that synthetic attribution must be made with more than a mere probability for the purposes of administrative convenience which can then withstand scrutiny.

Prong 3: Intent-Impact Calibration (I) - The Democratic Harm Gradient

The third one offers a consequential, adaptive response model driven by the relationship among measurable democratic effect of speech and intent to create or disseminate deepfakes. Mirroring a deepfake vs. non-deepfake case, not all deepfakes are treated equally under the law. A clear parody, clearly marked as such and intended to make a social point is at one end of an undemocratic harm gradient and at the other end is an unmarked deepfake film of a

political leader proclaiming election fraud that has gone viral through partisan networks 72 hours before polling. Restrictions have a firm constitutional footing.

Intent-Impact Calibration identifies three zones on the gradient of democratic harm, Deepfakes with clear indicators of parody, satire, or artistic intent but not much likelihood of being misunderstood as legitimate political communication are Zone 1. All Zone 1 deepfakes (other than those subject to technical disclosure requirements) are generally out of reach from regulators and attract the greatest level of Article 19(1)(a) protection. Zone 2 categorizes definitively harmful deepfakes including disinformation of ambiguous or contested intent, for example financially motivated false content, politically partisan propaganda without claim of synthetic creation, and user-generated video that is not materially manipulated but has an extensive distribution footprint. Zone 2 deepfakes are not illegal, but they can be flagged at the platform level and must be carefully labeled. Zone 3 consists of deceptive deepfakes that are intentionally produced to mislead voters, circulated covertly, and possess the potential to reach enough people so as to influence voting behaviour at an election. All types of regulatory action including mandatory removal by platforms, civil electoral fines and even unlimited criminal liability if culpability can be established apply to Zone 3 deepfakes.

This stepwise approach includes the democratic harm analysis developed in Sections IV and V, fitting within Shreya Singhal's proportionality imperative. It steers clear of the all or nothing choice between blanket protection and a complete prohibition that no democratic theory or constitutional doctrine could legitimately endorse.

Prong 4: Democratic Discourse Proportionality (D) - The Systemic Necessity Test

Importantly, the fourth pillar states that each regulatory tool must be as unrestrictive as possible in achieving the objective of democratic safeguard. This is ShreyaSinghal's famous strictosensu proportionality requirement restated, to the extent it succeeds in representing deepfake harms as systemic. The usual proportionality analysis asks whether a restriction on some particular instance of speech is needed to prevent the specific type of harm. To evaluate systemic deepfake harm, the test should ask whether, as a whole, the existing regulatory architecture is on balance the least restrictive means of protecting against epistemic harms to democratic deliberation at the level of scale where those harms are being felt.
The Democratic Discourse Proportionality testing consists of three parts. First, it requires a "use case at scale" study: the metric has to be tuned to the distributional structure of synthetic media harm instead of individual content pieces. This leads to a preference for upstream over downstream content prosecution as the more effective regulatory strategy, since upstream obligations do not have such severe side-effects on individual speech while adequately addressing systemic ills. Second, it requires an inquiry into a "democratic remainder": the measure should permit sufficient epistemic space for opposition discourse, satire, critique and genuinely adversarial political speech. However, if it means a regulation designed to tackle really bad information ends up silencing (as a side effect) legitimate political opposition by

inhibiting the production of satirical deepfakes, then all the democratic remaining can do is fail.Third, we need a "review mechanism" investigation into the measure's scope as deepfake production and detection technology progresses. A regulation that resembled a good fit for the technical environment of 2024 might be over- or under-inclusive by 2028; dynamic assessment rather than simply static design would be needed to maintain constitutional legitimacy.

VIII. Institutional Architecture: Operationalising the TRID Framework

Implementing the four prongs of the TRID Framework is dependent on an institutional architecture that can bring with it the constitutional accountability, procedural legitimacy, and technical capability that each prong requires. In this section, we describe three institutional improvements: an amendment to the Representation of the People Act, 1951; a Deepfake Adjudication Panel; and lastly, a Deepfake Electoral Integrity Protocol.

A. The Deepfake Adjudication Panel

To comply, in line with the Prong 2 Reasonable Attribution Standard this piece recommends a DAP be created to serve as a specialized, quasi-judicial body of record for regulatory activities concerning deepfakes. The three member DAP will consist of a retired High Court judge (Chairperson), a technical expert in computational media forensics and a civil society representative with experience in digital rights and electoral integrity, as it will be formed under The Information Technology Act, 2000 amended. This technical makeup of the DAP, along with the Supreme Court's insistence in Justice Puttaswamy on institutional safeguards commensurately suited to account for the "nature of the data and the harm" (paragraph 93), suggest that provision of forensic expertise at this stage in legal proceedings is constitutionally relevant more so at this investigatory-adjudicative interface than it may be merely at an investigatory threshold.
In Prong 2 applications, before regulatory action is triggered, the DAP would review attribution findings from both the Cyber Crimes Wing and the Election Commission; adjudicate interim removal orders in cases of electoral emergency via a 48-hour expedited procedure; maintain a public register of adjudicated deepfakes and applied forensic standards; conduct annual reviews on how deepfake-specific legislation should be restricted as part of Prong 4"review mechanism "element. The High Court would then be able to hear an appeal of the DAP's decisions under Article 226.

B. The Deepfake Electoral Integrity Protocol

DEIP (Deepfake Electoral Integrity Protocol) This article proposes a protocol, DEIP which will be implemented by the Election Commission of India within its constitutional authority under Article 324. The DEIP would operate as an electoral-period insertion into the Model Code of Conduct, giving certain responsibilities during the 60 days before any general or state legislative election. The DEIP would consist of four operational components: platform-level

obligations applying probabilistic deepfake detection to all electoral advertising materials, with periodic external audit for false negative rates; a dedicated Synthetic Media Rapid Response Cell within the Election Commission with the requisite technical forensic capability; and mandatory Content Authenticity Initiative (CAI) metadata requirements across digital platforms for all paid political advertising; and mandatory disclosure labels for AI-affiliated or synthetic content in political ads corresponding to the Article 50(4) thresholds of the EU AI Act, but with adjustments for India's unique electoral and constitutional context.
The DEIP would reduce the burden on downstream criminal enforcement by successfully targeting the concrete electoral harm in Prong 3 Zone 3 of the TRID Framework and by developing upstream detection, disclosure, and deterrence mechanisms that backloading content while reducing the volume of deepfake electoral content that is delivered to voters.

C. Proposed Amendment to the Representation of the People Act, 1951

This article Corrupt practice through synthetic media — a case for amending Section 123A of the Representation of the People Act, 1951 first appeared on AamJanata. Meanwhile, The "creation, distribution or financing of these synthetic electoral Media"—which is defined as audio-visual content that represents an electoral candidate, party official, or election authority that (a) has been materially generated or modified by computational means; (b) has not been explicitly disclosed at the point of original publication as being Synthetic; and (c) was created with the intent to manufacture prejudice toward a voter preference on behalf of one or more candidates for office would become subject to a When the Content attained proven mass propagation — in excess of 250K views-- accountability would be enforced viciously against its dissemination (exceeding 100,000 views) within 30 days of the polling date.
Constitutionally, it is sustained by the longstanding underpinning of decency in political campaigns as part of the Representation of the People Act, Article 19(2) justifications over and above that could be constitutionally extended for "sovereignty and integrity of India," and what gives electoral law a unique power to regulate conduct standards for political actors which you would not have otherwise been able to impose through content-neutral speech restrictions outside an election context - this all backs up Section 123A. A structured proportionality analysis done within the TRID Framework supports constitutional adequacy for such a provision, especially considering the Supreme Court ruling in Subramanian Swamy v Election Commission of India ((2013) 10 SCC 500) affirmed Parliament's wide latitude to legislate against those things that affect election purity.

IX. Conclusion: Reconstituting Article 19(1)(a) for the Deepfake Era

As we argue in this article, deepfake-driven deception or "epistemic sabotage" represents an entirely novel constitutional threat that requires not only a new institutional architecture but also a new doctrinal standard. Currently, while the Information Technology Act, DPDP Act along with Intermediary Guidelines and electoral legislation exist, the regulatory environment is insufficient- it lacks other systemic architecture because attribution isn't possible without a

strong technical forensic capacity governing such online speech trends; yet individual cases tend to be looked at more broadly but curbs are only pre-emptively implemented through upstream deterrents or remedies imposed post hoc. This comparative research demonstrates that even more effective and constitutionally compliant regulatory approaches are available.

This article proposes a new TRID Framework that works through the Threshold of Harm, Reasonable Attribution Standard, Intent-Impact Calibration and Democratic Discourse Proportionality to give legislators and courts an implementable constitutional standard that will leave intact the very values like political contestation, pluralism, inventiveness that Article 19(1)(a) sets out to protect. One of the Frameworks main advantages is its doctrine specific application, which can provide reproducible results in limited cases, as well as agile for change through integrated review processes. Absent the enforcement structure provided by its institutional complements the Deepfake Adjudication Panel, the Deepfake Electoral Integrity Protocol (in this case, codified by law), and the proposed Section 123A(iii) any doctrinal standard would be aspirational.

The more profound constitutional argument contained in this article, which goes beyond rhetorical flourish, is that Article 19(1)(a) derives momentum from being the product of a democracy. This is a threshold commitment: that the Constitution's protection for free expression serves to promote democratic self-governance, not simply individual communicative autonomy as an end in its own right. Article 19(1)(a) therefore necessitates regulation rather than mere permission of synthetic media when such practices systematically unfound the epistemic preconditions for democratic self-governance. In a technological age, India has no choice but to build the constitutional functionalities that must safeguard both democratic integrity and free speech—this is the challenge for any legislature and courts of law in India. To assist in this endeavor, we offer the TRID Framework.

References

- Alkiviadis, P. (2021). Regulating deepfakes: Between information freedom and epistemic harm. European Journal of Law and Technology, 12(2), 1-28.
- AnuradhaBhasin v. Union of India, (2020) 3 SCC 637 (Supreme Court of India).
- Bakir, V., &McStay, A. (2018).Fake news and the economy of emotions. Digital Journalism, 6(2), 154-175. https://doi.org/10.1080/21670811.2017.1345645
- Balkin, J. M. (2018). Free speech in the algorithmic society: Big data, private governance, and the new botnet politics. UC Davis Law Review, 51(3), 1149-1210.
- Bennett Coleman & Co. v. Union of India, (1973) 2 SCC 788 (Supreme Court of India).
- Bhatia, G. (2016). Offend, shock, or disturb: Free speech under the Indian Constitution. Oxford University Press.
- Brandenburg v. Ohio, 395 U.S. 444 (1969) (Supreme Court of the United States).
- California Assembly Bill 730. (2019). Elections: deceptive audio or visual media. State of California.
- Chander, A., & Le, U. P. (2020).Data nationalism. Emory Law Journal, 64(3), 677-739.
- Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107, 1753-1820.

- Citron, D. K., & Chesney, R. (2019).Deep fakes and the new disinformation war. Foreign Affairs, 98(1), 147-155.
- Deibert, R. J. (2019). The road to digital unfreedom: Three painful truths about social media. Journal of Democracy, 30(1), 25-39. https://doi.org/10.1353/jod.2019.0002
- Election Commission of India.(2019). Model Code of Conduct for the guidance of political parties and candidates.Election Commission of India.
- European Commission. (2022). Regulation (EU) 2022/2065 on a Single Market for Digital Services (Digital Services Act). Official Journal of the European Union, L 277, 1-102.
- European Commission. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.
- Farid, H. (2022). Creating, using, misusing, and detecting deep fakes. Journal of Online Trust and Safety, 1(4), 1-18. https://doi.org/10.54501/jots.v1i4.56
- Guess, A. M., & Lyons, B. A. (2020). Misinformation, disinformation, and online propaganda. In N. Persily& J. A. Tucker (Eds.), Social media and democracy: The state of the field, prospects for reform (pp. 10-33). Cambridge University Press.
- Habermas, J. (1989). The structural transformation of the public sphere: An inquiry into a category of bourgeois society (T. Burger, Trans.). MIT Press.
- Indian Express Newspapers (Bombay) Pvt. Ltd. v. Union of India, AIR 1986 SC 515 (Supreme Court of India).
- Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1 (Supreme Court of India).
- Krishnaswamy, S. (2009). Democracy and constitutionalism in India: A study of the basic structure doctrine. Oxford University Press.
- Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. Journal of Applied Research in Memory and Cognition, 6(4), 353-369. https://doi.org/10.1016/j.jarmac.2017.07.008
- Michaels, R. (2006). The functional method of comparative law.In M. Reimann& R. Zimmermann (Eds.), The Oxford handbook of comparative law (pp. 339-382).Oxford University Press.
- Ministry of Electronics and Information Technology. (2021). The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. Gazette of India.
- Ministry of Electronics and Information Technology.(2023). Digital Personal Data Protection Act, 2023.Gazette of India, No. 60.
- New York Times Co. v. Sullivan, 376 U.S. 254 (1964) (Supreme Court of the United States).
- Paris, B., & Donovan, J. (2019).Deepfakes and cheap fakes: The manipulation of audio and visual evidence. Data & Society Research Institute.
- Rao, A. (2021). Surveillance, privacy, and the Indian state: Toward a critical framework. National Law School of India Review, 33(1), 1-38.
- Rawls, J. (1993). Political liberalism.Columbia University Press.
- Rini, R. (2017). Fake news and partisan epistemology.Kennedy Institute of Ethics Journal, 27(2), E-43-E-64. https://doi.org/10.1353/ken.2017.0025
- RomeshThappar v. State of Madras, AIR 1950 SC 124 (Supreme Court of India).
- S. Rangarajan v. P. Jagjivan Ram, (1989) 2 SCC 574 (Supreme Court of India).
- Sakal Papers (P) Ltd. v. Union of India, AIR 1962 SC 305 (Supreme Court of India).

- Secretary, Ministry of Information and Broadcasting v. Cricket Association of Bengal, (1995) 2 SCC 161 (Supreme Court of India).
- ShreyaSinghal v. Union of India, (2015) 5 SCC 1 (Supreme Court of India).
- Simons, G. (2019). Fake news: A challenge for democracy. Strategic Analysis, 43(1), 55-64. https://doi.org/10.1080/09700161.2019.1573129
- Sinha, G. (2020). Intermediary liability in India: The 2021 Rules and their constitutional implications. Journal of Indian Law and Society, 12(2), 45-89.
- Subramanian Swamy v. Election Commission of India, (2013) 10 SCC 500 (Supreme Court of India).
- Sunstein, C. R. (2017). #Republic: Divided democracy in the age of social media. Princeton University Press.
- Susskind, J. (2022). The digital republic: On freedom and democracy in the 21st century. Bloomsbury Publishing.
- Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020).Deepfakes and beyond: A survey of face manipulation and fake detection. Information Fusion, 64, 131-148. https://doi.org/10.1016/j.inffus.2020.06.014
- United Kingdom Parliament.(2023). Online Safety Act 2023.His Majesty's Stationery Office.
- United States v. Alvarez, 567 U.S. 709 (2012) (Supreme Court of the United States).
- Wardle, C., &Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe.
- Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.