Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Deepfakes and Democracy: Reputation, Disinformation, and Freedom of Expression in the Age of Synthetic Media

Alok Debnath[1], Dr. Ajaz Afzal Lone[2]

[1]LL.M. Student, University Institute of Legal Studies, Chandigarh University, Mohali, Punjab, India
[2]Associate Professor, University Institute of Legal Studies, Chandigarh University, Mohali, Punjab, India.

DOI: https://doi.org/10.63363/aijfr.2026.v07i06.8971

Abstract

Deepfakes, synthetic audio, images and video that convincingly portray people saying and/or doing things they didn't do and/or say, have become common tools for fraud, harassment and political manipulation.[1] They pitted two democratic commodities against one another. The right to freedom of expression also covers speech that one does not want or may not even agree with, in part because, through open discussion, free societies check ideas and hold power accountable. The right to reputation is concerned with the dignity of the individual and his or her reputation or standing, and also indirectly the integrity of the information for voters. The paper analyses how that conflict is being resolved in India, in the U.S. and the European Union, and considers the principles of the constitution, as well as the regulatory and judicial landscape as of late 2026. These encompass India's 2026 amendments to the IT Rules,[2] EU AI Act's Article 50 transparency regime,[3] the US TAKE IT DOWN Act,[4] the Kohls v. Bonta litigation on the California election deepfake statutes[5] and recent Delhi High Court rulings on personality rights and political satire.[6] The paper concludes that a blanket prohibition and regulatory abstention are not justifiable. It is better to have four tiers of this: disclosure is the standard remedy for synthetic content; narrowly drafted liability

[1] *See* Henry Ajder, Giorgio Patrini, Francesco Cavalli & Laurence Cullen, *The State of Deepfakes: Landscape, Threats, and Impact* (Deeptrace, Sept. 2019) [hereinafter Ajder *et al.*].

[2] The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, G.S.R. 139(E) (25 Feb. 2021), as amended by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 (notified 10 Feb. 2026) [hereinafter IT Rules].

[3] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), 2024 O.J. (L) 2024/1689, art. 50 [hereinafter AI Act].

[4] Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act (TAKE IT DOWN Act), Pub. L. No. 119-12, 139 Stat. 55 (2025) [hereinafter TAKE IT DOWN Act].

[5] *Kohls v. Bonta*, No. 2:24-cv-02527 (E.D. Cal. 2024–2025) [hereinafter *Kohls*]; Assembly Bill No. 2839 and Assembly Bill No. 2655 (2023–24 Reg. Sess.) (Cal. 2024).

[6] *Raghav Chadha v. Ashok Kumar (John Doe) & Ors.*, CS(OS) 466/2026 (Delhi High Court, 1 July 2026) [hereinafter *Raghav Chadha*]; *Yuvraj Singh v. Ashok Kumar & Ors.*, CS(COMM) 796/2026 (Delhi High Court, 29 July 2026) [hereinafter *Yuvraj Singh*].

for knowingly deceptive depictions in situations of concrete injury; speedy but procedurally protected remedies for the worst harms, like non-consensual intimate imagery; and substantial protection for satire and political criticism.

Keywords: deepfakes; synthetic media; freedom of speech; right to reputation; defamation; personality rights;

1. Introduction

An audio recording emerged online recently in the closing days leading up to Slovakia's parliamentary election this year in 2023, apparently while a prominent candidate was debriefing on how to rig the vote.[7] It was claimed to be artificially produced. In January 2024, an artificial voice with the likeness of President Joe Biden called voters in New Hampshire to dissuade them from voting in a primary.[8] Both episodes didn't take much in the way of high-tech equipment. Both demonstrated that a convincing forgery can be created at low cost, at scale, and at the time of greatest need to correct the error.

The legal issue is as old as it is new. In the past, forged letters, doctored photographs and false rumours were a problem for law, as were the laws that had to balance the harm they cause with the value of free expression. The essential difference between deepfakes is that the falsehood is now obviously fake. When the speaker makes a written lie, he/she is asking the audience to trust him/her. A fake video is going to ask people to believe what they see and hear, and it's a challenge that human beings are not up to.[9] Meanwhile, the same technology that creates a politician's confession can create a satirical sketch, an accessibility tool, a film restoration or a political cartoon in motion. Any law that is general enough to enjoin the first will do the same for the second.

This paper asks a particular question: What is the balance of the freedom of speech and right to reputation when the speech is synthetic instead of real? It goes through a few stages. In Part 2, the technology and the harms it causes to individuals and to democratic processes are described. Part 3 describes the two conflicting rights conceptually. Part 4 looks at the constitutional and doctrinal structures in India, United States and Europe. In Part 5, the latest regulatory reactions are examined. Part 6 analyses three recent developments as case studies. Part 7 recommends a framework for principled balancing, Part 8 highlights unaddressed challenges, and Part 9 presents recommendations.

There are three methodological notes. First, this paper is comparative and doctrinal in nature, and not based on original empirical research. Secondly, this is a fast-evolving area of law and some of the changes mentioned are only of recent vintage - and some are based on secondary reporting of cases, so in such cases they are identified as such and should be checked against primary sources prior to any submission. Third, the paper starts by considering "reputation" in a broad sense, covering not only defamation, but also dignity, privacy and the personality or publicity rights, which courts in recent years have made greater use of against synthetic impersonation, especially in India.[10]

[7] *See* Morgan Meaker, "Slovakia's Election Deepfakes Show AI Is a Danger to Democracy", *Wired* (3 Oct. 2023).

[8] Federal Communications Commission, Declaratory Ruling, *Implications of Artificial Intelligence Technologies on Protecting Consumers from Unwanted Robocalls and Robotexts*, CG Docket No. 23-362, FCC 24-17 (8 Feb. 2024).

[9] Robert Chesney & Danielle Keats Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", 107 Cal. L. Rev. 1753 (2019) [hereinafter Chesney & Citron].

[10] *D.M. Entertainment Pvt. Ltd. v. Baby Gift House*, 2010 (44) PTC 429 (Del) [hereinafter *D.M. Entertainment*]; *Anil Kapoor v. Simply Life India*, 2023 SCC OnLine Del 6914 [hereinafter *Anil Kapoor*]; *see infra* Part 4.1.

2. Understanding Deepfakes: Technology, Uses and Harms

2.1 What a deepfake is

Deep fake" is a portmanteau of the words "deep learning" and "fake." Media created or edited by a machine-learning system, usually a generative adversarial network or diffusion and transformer-based models,[11] to look like a real person's face, body or voice doing or saying something that never actually happened. There is no consensus on the definition among legislators. The EU AI Act's definition of the deepfake is based on AI-generated or manipulated content that looks like an existing person, object, place or event and would seem real to a person.[12] The changes to the IT Rules in India are not limited to the word, but also include changes to the definition of synthetically generated information, which, by substance, includes audio, visual or audio-visual content produced artificially or algorithmically in a manner that makes it appear real.[13] Both definitions hinge on the perceived authenticity, which also has a legal dimension: a cartoonish depiction that a reasonable viewer would not think to be truthful is very different from a photo-realistic forgery.

2.2 Legitimate and expressive uses

The first step in any serious legal analysis of synthetic media is to recognize that synthetic media is not a bad thing. It is used in film dubbing and de-ageing, to give voice to those who have lost their vocal ability, to educate and present in museums, to parody and satire, and in political commentary in the long tradition of caricature. If you have a rule that affects the technology and doesn't affect the deception, then you sweep in all of these uses. That is why courts that enforce free-speech guarantees repeatedly have turned to the question whether a statute makes a distinction between deceptive and non-deceptive use.[14]

2.3 Harms to individuals

The individual harms of deepfakes fall into at least four overlapping categories.

- Sexual abuse and intimate abuse- Research on the deepfake industry, dating back to the first reports, has consistently revealed that a significant portion of deepfakes online is pornographic, and that women are the predominant victims.[15] These are those expressing little or no legitimate function, and causing serious and irreversible damage to dignity, mental health and livelihood.
- Reputational defamation- The image of a man confessing to a crime, supporting a product or the expression of a slur have the same classic injury to reputation, and the added element is that it is hard to show in a rebuttal once the viewers have "seen" the event.
- Fraud and impersonation- The voice and video cloning is employed to impersonate family members, executives or officials and extract money and credentials. The damage is mainly monetary but also involves the exploitation and loss of trust of a person's identity.

[11] Ian J. Goodfellow *et al.*, "Generative Adversarial Nets", in 27 *Advances in Neural Information Processing Systems* 2672 (2014).

[12] AI Act, *supra* note 3, art. 3(60).

[13] IT Rules, *supra* note 2 (definition of "synthetically generated information").

[14] *United States v. Alvarez*, 567 U.S. 709 (2012) [hereinafter *Alvarez*]; *Kohls*, *supra* note 5.

[15] Ajder *et al.*, *supra* note 1.

- Commercial misappropriation- Celebrities and public figures have their likeness and voice exploited without their permission or knowledge to sell products or to promote scams, a trick that falls at the intersection of reputation and property rights, and over which Indian courts have offered some oversight via the jurisprudence of personhood rights.[16]

2.4 Harms to democratic processes

The damage to democracy is more difficult to quantify and more hotly debated. There are three mechanisms that are generally recognized.

Direct deception- Voters may not have time to check such a video or recording if it's created just before a poll, and it can deceive them. The typical examples are the Slovak incident or the incident in New Hampshire.[17] There was mixed verification for the initial big wave of elections in 2024. A number of analyses found little evidence for synthetic media having a significant impact on election outcomes,[18] but also noted that it was widely used for harassment, to mock opponents and to reinforce existing partisan narratives. A sensible perspective would be that there is a risk but the size of the risk is uncertain and that is important, because the more uncertain the risk, the more tempting and the more costly it is to speech over-regulate.

Abuse of participants, targeted abuse- Sexualised deepfakes have been disproportionately used to humiliate or oust women candidates and journalists from public life. Not only is it a personal harm, but it is also a democratic one because it cuts down on who can run for office or go in front of an audience.

The liar's dividend- It's a scenario that has been raised by scholars, such as Robert Chesney and Danielle Citron,[19] who have pointed out that once the public is aware of convincing fakes, real recordings can be ignored as fake. If the politician himself is captured on real tape, he can say it is a deep fake. This second-order effect is that deepfakes undermine democratic conversation even if they are not specifically believed to be fake, as they reduce the epistemic value of audio and video evidence in general.

3. Conceptual Foundations: Two Rights in Tension

3.1 Why free expression protects falsehood

Reasons for the protection of speech are also well known: the need to seek truth by free competition of ideas,[20] the need to have the liberty to act upon one's own conscience and exercise one's own free will, and the democratic one that if people are unable to govern themselves because of what they are allowed to hear, they will not be free.[21] In the context of an election, the democratic argument carries extra weight, and the democratic argument is the one that is politically appealing in the context of deepfake regulation. In practice, governments that control "false" political speech determine what is "false," and every government official has an incentive to say that something that is inconvenient to them is "disinformation."

[16] *D.M. Entertainment*, *supra* note 10; *Anil Kapoor*, *supra* note 10; *see infra* Part 4.1.

[17] *See supra* notes 7, 8.

[18] *See, e.g.*, Sam Stockwell *et al.*, *AI-Enabled Influence Operations: Safeguarding Future Elections* (Centre for Emerging Technology and Security, The Alan Turing Institute, Nov. 2024).

[19] Chesney & Citron, *supra* note 9.

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

[21] *See Shreya Singhal v. Union of India*, (2015) 5 SCC 1 : AIR 2015 SC 1523 [hereinafter *Shreya Singhal*].

There is also one more aspect more specific to this topic. Courts have been hesitant to acknowledge any broad and undefined principle of "false speech" that is not protected. In United States v. Alvarez (2012), the United States Supreme Court upheld an anti-doublespeak statute against constitutional challenges, stating that lies are not necessarily outside the First Amendment protections.[22] However, lies that cause legally cognisable harm like defamation and fraud are subject to regulation.[23] This reasoning has direct implications for statutes against deepfakes: a law to ban "materially deceptive" media but not actual damage will have a hard time to prove it.

3.2 Why reputation deserves legal protection

In international and national law, reputation is safeguarded as a part of human dignity and the private sphere. Under Article 17 of the International Covenant on Civil and Political Rights, any attack on honour or reputation is illegal, while under Article 19(3) of the same Covenant restrictions on expression are acceptable when they are 'necessary to secure the enjoyment by others [of the rights guaranteed in the Covenant]'.[24] The European Convention on Human Rights (ECHR) has been understood to encompass the right to private life, including the right to reputation, in Article 8 and there is a specific exception for restrictions that are necessary for the protection of the reputation or rights of others in Article 10(2).[25] In India, the Supreme Court has ruled that reputation is a part of Article 21 of the Indian Constitution (Right to life and personal liberty)[26] and that criminal defamation is a reasonable restriction on Article 19(2) of the Constitution (Right to free speech).[27]

Reputation also is instrumental in a democratic usage. Without the voters' confidence that the statements and images that are associated with candidates are genuine, informed choice is impossible. In this context, the right to reputation is not just the private interest of the portrayed person, but also a precondition for the information to be received by the public in an accurate way.

3.3 Why deepfakes unsettle the traditional balance

Traditional rules for reconciliation of speech/reputation arose in a world of reading, listening, and viewing. These are the three ways that deepfakes challenge these doctrines.

First **the factual-assertion test-** Most jurisdictions uphold freedom of opinion, rhetorical exaggeration and satire, and only hold for liability false statements of fact which a reasonable person would expect to be literal.[28] Deepfakes make this distinction less clear-cut. A photorealistic video is perceived as a record of fact by default — the context or labels that indicate parody in a cartoon may not be present or removed when the content is passed on.

[22] *Alvarez*, *supra* note 14 (plurality opinion of Kennedy, J.) (invalidating the Stolen Valor Act of 2005, 18 U.S.C. § 704(b)).

[23] *Ibid.*

[24] International Covenant on Civil and Political Rights, 16 Dec. 1966, 999 U.N.T.S. 171, arts. 17, 19(3) [hereinafter ICCPR].

[25] Convention for the Protection of Human Rights and Fundamental Freedoms, 4 Nov. 1950, 213 U.N.T.S. 221, arts. 8, 10(2) [hereinafter ECHR].

[26] *Subramanian Swamy v. Union of India*, (2016) 7 SCC 221: AIR 2016 SC 2728 [hereinafter *Subramanian Swamy*]; The Constitution of India, 1950, art. 21.

[27] The Constitution of India, 1950, arts. 19(1)(a), 19(2).

[28] *Hustler Magazine, Inc. v. Falwell*, 485 U.S. 46 (1988) [hereinafter *Hustler*]; *see infra*-Part 4.2.

Secondly, **scale and speed-** Litigation will ensue after publication—often years later—is the assumption of defamation law. This is a viral deep fake that can cause harm in hours. This drives regulators to take action on the admin side,[29] and speeds up the possibility of a mistake or abuse.
Thirdly, **anonymity and provenance-** Creators are often anonymous or overseas users, and the point of control has become intermediaries and platforms.[30] What is the role of a platform, and how quickly, what kind of notice, of "may this speaker be punished?" moving to "what must a platform do?" The answer will dictate how much of lawful speech is taken away by overly precautionary compliance.

4. Constitutional and Doctrinal Frameworks

4.1 India

The basic principles of the Indian freedom of speech law are enshrined in Article 19(1)(a) as well as the eight grounds of restriction under Article 19(2), including defamation.[31] In Subramanian Swamy v. Union of India (2016), the Supreme Court had indeed upheld the constitutionality of criminal defamation,[32] basing its judgment on the fact that reputation is a part of the right to life with dignity as enshrined in Article 21 of the Constitution and the restriction is reasonable. The concept of criminal defamation has been introduced in the Bharatiya Nyaya Sanhita, 2023 in section 356.[33] The balance is then, in formal teaching, more on the reputation side than in the United States.
Concurrently, the ruling in Shreya Singhal v. Union of India[34] from the Court is a key protection against over-reach in the digital world. It found section 66A of the Information Technology Act[35] to be vague and chilling because it failed to clarify the definition of a court order or government notification under the safe harbour provision; and section 79 of the Information Technology Act[36] to be susceptible to reading-down because it allows the safe harbour to be breached where the platform has "actual knowledge" of the infringement, which, in the absence of any further statutory definition, could be interpreted to mean that it has been notified by a court or government. That interim injunctions against publication are exceptional are reinforced in subsequent decisions. In Bloomberg Television Production Services India Pvt. Ltd. v. Zee Entertainment Enterprises Ltd. (2024),[37] the Supreme Court, following the English precedent of Bonnard v. Perryman,[38] has held that gag orders should be avoided in defamation cases prior to trial, as premature imposition of such restrictions can stifle legitimate public discourse.[39]

[29] *See infra* Part 5.1.
[30] The Information Technology Act, 2000 (Act 21 of 2000), s. 79 [hereinafter IT Act].
[31] The Constitution of India, 1950, arts. 19(1)(a), 19(2).
[32] *Subramanian Swamy*, *supra* note 26 (upholding the Indian Penal Code, 1860, ss. 499–500 and the Code of Criminal Procedure, 1973, s. 199).
[33] The Bharatiya Nyaya Sanhita, 2023 (Act 45 of 2023), s. 356 [hereinafter BNS].
[34] *Shreya Singhal*, *supra* note 21.
[35] IT Act, *supra* note 30, s. 66A (struck down in *Shreya Singhal*).
[36] IT Act, *supra* note 30, s. 79(3)(b); *Shreya Singhal*, *supra* note 21 (reading down s. 79(3)(b) and r. 3(4) of the Information Technology (Intermediaries Guidelines) Rules, 2011).
[37] *Bloomberg Television Production Services India Pvt. Ltd. v. Zee Entertainment Enterprises Ltd.*, 2024 SCC OnLine SC 426 (Supreme Court of India, 22 Mar. 2024) [hereinafter *Bloomberg*].
[38] *Bonnard v. Perryman*, [1891] 2 Ch 269 (CA) [hereinafter *Bonnard*].
[39] *Bloomberg*, *supra* note 37.

There are two additional lines of authority. The Court's judicial endorsement of the concept of privacy as a fundamental right in K.S. Puttaswamy v. Union of India (2017) gives a constitutional basis for asserting a right to be free from non-consensual synthetic depiction, especially intimate imagery.[40] The principle that public officers have to tolerate more scrutiny and criticism than private persons are supported by the case of And R. Rajagopal v. State of Tamil Nadu (1994).[41]

There is also a parallel development in the Delhi High Court which has been underlying since D.M. Entertainment v. Baby Gift House (2010),[42] where the court has laid great emphasis on protecting a performer's name and likeness from unauthorised commercial use. In 2025 and 2026, that court has issued a steady stream of interim orders, often against a faceless “John Doe” defendant, that prohibit the use of the name, image, voice and likeness of film actors, cricketers, spiritual leaders, journalists and entrepreneurs in the creation of deepfakes and unauthorised merchandise.[43] These orders usually require a quick removal of identified contents from the platforms. Their reach is impressive and so is their composition: the recipients are overwhelmingly familiar and well-resourced; this will be discussed again in Part 8.[44]

4.2 The United States

This is the most speech-protective position of the three from the American. In New York Times Co. v. Sullivan (1964), a public official cannot recover for defamation unless he establishes actual malice (actual knowledge that the statement was false or reckless disregard of that fact).[45] In Hustler Magazine v. Falwell (1988), the Court extended this protection to parody: to be recoverable under the intentional infliction of emotional distress doctrine, the publication of a parody must contain a false statement of facts that could be interpreted as a factual statement.[46] In these rulings, American courts have a test for deepfakes already on hand. A realistic, deceptive and knowingly false may be actionable defamation, but only if the usual elements are satisfied.

There are three additional aspects of the American perspective. Second, the freedom of the press and the media is inextricably linked to the freedom of the press and the media. Second, the freedom of the press and the media is inextricably linked to the freedom of the press and the media. Second, platforms are protected from liability under Section 230 of the Communications Decency Act[47] for the majority of user-generated content, which has meant that regulatory pressure has had to be exercised in other ways, including the notice-and-takedown scheme of the TAKE IT DOWN Act implemented by the Federal Trade Commission.[48] Thirdly, a number of states have their own deepfake laws, including some on sexual

[40] *Justice K.S. Puttaswamy (Retd.) v. Union of India*, (2017) 10 SCC 1: AIR 2017 SC 4161 [hereinafter *Puttaswamy*].
[41] *R. Rajagopal v. State of Tamil Nadu*, (1994) 6 SCC 632: AIR 1995 SC 264 [hereinafter *R. Rajagopal*].
[42] *D.M. Entertainment*, *supra* note 10.
[43] *Yuvraj Singh*, *supra* note 6; *Anil Kapoor*, *supra* note 10.
[44] *See infra* Part 8.
[45] *New York Times Co. v. Sullivan*, 376 U.S. 254, 279–80 (1964) [hereinafter *Sullivan*].
[46] *Hustler*, *supra* note 28.
[47] 47 U.S.C. § 230(c)(1).
[48] TAKE IT DOWN Act, *supra* note 4, § 3.

imagery, election messages, and the commercial exploitation of a person's image via right of publicity.[49] The election statutes have been most easily open to constitutional attack and Part 6 explains why.[50]

4.3 Europe

The European law perspective of the conflict is more of a proportionality one than a categorical one. The European Convention on Human Rights (ECHR) protects expression, including speech which offends, shocks or disturbs,[51] and protects private life and, as the Strasbourg Court has said, reputation.[52] The criteria used to balance the two are outlined in the Grand Chamber of the European Court of Human Rights in the case of Axel Springer AG v. Germany (2012):[53] the contribution to a debate of general interest, the extent to which the person in question is known and the subject of the report, the other actions of the person concerned, the means used to obtain the information and their reliability, the content, form and consequences of the publication, and the gravity of any sanction.

These criteria are applicable to deepfakes. The factors of "veracity" and "method" predominate over the fabricated depictions, and the factors of "general interest" and "form" protect satire and commentary. The EU law builds on this through provisions on data protection (e.g. the face and voice of a person are personal data),[54] the Digital Services Act, which requires very large platforms to evaluate and reduce systemic risks, such as threats to electoral processes[55]—and, most recently, the transparency regime of the AI Act (Part 5). The European solution is therefore a layered, procedural one, based on the principle of disclosure and risk assessment, with a proportionality review on a case-by-case basis.

5. Recent Regulatory Responses

5.1 India: the 2026 amendments to the IT Rules

The most important statutory measure taken by India is Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 notified by the Ministry of Electronics and Information Technology on 10 February 2026 and its implementation from 20 February 2026.[56] The amendments add a new category of "synthetically generated information," which is understood to mean audio, visual or audio-visual information that is artificially or algorithmically created or modified to seem real. The definition does not include "text-only" content. But it will require intermediaries to make sure that the content is clearly marked, that it is annotated with provenance metadata, and that users explicitly assert whether content is synthetic or not, the platforms are expected to implement reasonable and proportionate technical measures to confirm these assertions. The amendments also define "information" in the intermediary rules, including to cover synthetic information, and also expand on the secondary rules

[49] *See, e.g.*, Cal. Elec. Code § 20012; N.Y. Civ. Rights Law §§ 50–51.

[50] *See infra* Part 6.1; *Kohls*, *supra* note 5.

[51] *Handyside v. United Kingdom*, App. No. 5493/72 (ECtHR, 7 Dec. 1976), 1 EHRR 737.

[52] ECHR, *supra* note 25, arts. 8, 10; *Pfeifer v. Austria*, App. No. 12556/03 (ECtHR, 15 Nov. 2007).

[53] *Axel Springer AG v. Germany*, App. No. 39954/08 (ECtHR, Grand Chamber, 7 Feb. 2012): (2012) 55 EHRR 6 [hereinafter *Axel Springer*].

[54] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 Apr. 2016 (General Data Protection Regulation), 2016 O.J. (L 119) 1, art. 4(1) [hereinafter GDPR].

[55] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 Oct. 2022 on a Single Market for Digital Services (Digital Services Act), 2022 O.J. (L 277) 1, arts. 34–35 [hereinafter DSA].

[56] IT Rules, *supra* note 2.

to clarify sharply diminished takedown timelines, with a three-hour window for certain categories of harmful information.[57]

The amendments have two sides from a rights point of view. In principle, the tools that cause the least restriction on speech are mandatory labelling and provenance. But there are murky areas of intent and free speech, and pressure from the need to remove content before it can be assessed. The conflict between Shreya Singhal's desire to have the knowledge aspect supervised at the court level[58] ought to be followed up in litigation as it is tested as to the rules.

5.2 The European Union: Article 50 of the AI Act

Transparency is the main tool selected by the EU. Providers of generative systems have to label synthetic output as such in a way that is visible and machine-readable,[59] and the deployer of a deepfake has to disclose synthetic or manipulated content when publishing it, under Article 50 of the AI Act, which comes into effect on 2 August 2026.[60] For content that is part of an obviously artistic, creative, satirical or fictional piece, the disclosure requirement is mitigated to not destroy the piece, but information must be provided.[61] European guidelines and a code of practice on these duties were published in mid-2026[62] and a transition period will postpone the obligation to mark generative systems until 2 December 2026, if they were already available on the market before August. Breaches are grouped in the second tier of the penalties provided by the Act and could be fined up to €15 million or 3 per cent of an undertaking's worldwide annual turnover.[63]

The design is unique in what it doesn't do. It does not create a new cause of action for the subject of the picture, and does not bar the use of a "synthetic" depiction of a real person. It seeks to help the audience regain the ability to tell the difference between genuine and virtual content – questions of defamation and privacy will be addressed by current national and Union law.

5.3 The United States: TAKE IT DOWN and state election laws

The TAKE IT DOWN Act, signed in May 2025, mandates that covered platforms implement a removal procedure for those who are the victim of a non-consensual intimate image and can request that the image be removed and known duplicate images be removed within 48 hours of a valid request at the federal level.[64] The provision will go into effect on 19 May 2026 and the Federal Trade Commission has stated that it can impose civil penalties of more than $53,000 per violation.[65] Civil-liberties organizations have attacked the statute for failing to have a counter-notice system or punish falsified requests, because it could lead to over-removal of legal material, and its supporters have pointed out that it only applies to intimate imagery, which has little protected value.

[57] *Ibid.*

[58] *Shreya Singhal*, *supra* note 21 (reading downs. 79(3)(b) of the IT Act, *supra* note 30).

[59] AI Act, *supra* note 3, art. 50(2).

[60] AI Act, *supra* note 3, arts. 50(4), 113.

[61] *Ibid.*, art. 50(4).

[62] European Commission, guidelines and code of practice on transparency obligations under Article 50 of the AI Act (2026).

[63] AI Act, *supra* note 3, art. 99(4).

[64] TAKE IT DOWN Act, *supra* note 4, § 3.

[65] *Ibid.*; *see also* Federal Trade Commission, enforcement announcements (May 2026).

In the state this doesn't apply. There are laws in place in legislatures about synthetic media during elections and laws regarding the use of likeness for commercial purposes. There are laws in place in legislatures about synthetic media during elections and laws regarding the use of likeness for commercial purposes. Statutes that ban or require removal of misleading media have struck First Amendment challenges several times,[66] and, in the absence of final decisions, stating disclosure requirements, such as the model coming from Arizona in commentary, have done well.

5.4 Likeness rights as an alternative architecture

Another way of thinking about the situation is to see it as a struggle for control over one's identity instead of a quest for falsity. Denmark has pledged further legislation, described in the press as providing an "almost copyright" right of individuals over their face, body and voice, and allowing for the exceptions of parody and satire.[67] The bill passed the legislative stage and entered into effect at various dates as reported, and should be verified with official Danish sources before use as in effect. In India, the judiciary creates its own law, and does the same for personality rights.[68] The advantages of likeness models are that there is no need to prove falsity or reputational damage, and disadvantages is that they can be used to silence criticism or commentary that is simply a person's dislike of the depicted person, making a dignity interest akin to a 'property veto' over speech.

6. Three Case Studies

6.1 *Kohls v. Bonta*: when election deepfake laws meet the First Amendment

In 2024, California passed a suite of bills to combat election media deception. Assembly Bill 2839 limited the dissemination of “materially deceptive” election-related messaging in the period leading up to and after an election, while also enabling private lawsuits,[69] and Assembly Bill 2655 mandated that large platforms block and mark “materially deceptive” election-related messages.[70] A political satire video maker and X Corp. and a satirical media outlet sued. Judge John Mendez of the Eastern District of California issued an initial injunction against AB 2839 in October 2024,[71] and ruled in Favor of Section 230 against AB 2655 in August 2025.[72]Judge John Mendez of the Eastern District of California issued an initial injunction in October 2024 regarding AB 2839, and ruled in Favor of Section 230 against AB 2655 in August 2025. Based on published summaries of the reasoning,[73] the court determined that AB 2839 was a content- and viewpoint-based restriction which was not limited to the "least restrictive means,"[74] it was not within the defamation exception, and its legitimate sweep was small compared to its application to

[66] *Kohls*, *supra* note 5.

[67] Proposal to amend the Danish Copyright Act (*Ophavsretsloven*) to protect personal likeness (Denmark, 2025); *see* reporting in *The Guardian* and *Biometric Update*.

[68] *See supra* Part 4.1; *D.M. Entertainment*, *supra* note 10; *Anil Kapoor*, *supra* note 10.

[69] Assembly Bill No. 2839 (2023–24 Reg. Sess.) (Cal. 2024), *codified at* Cal. Elec. Code § 20012.

[70] Assembly Bill No. 2655 (2023–24 Reg. Sess.) (Cal. 2024).

[71] *Kohls*, *supra* note 5 (order granting preliminary injunction, Oct. 2024).

[72] *Kohls*, *supra* note 5 (Aug. 2025) (holding AB 2655 pre-empted by 47 U.S.C. § 230).

[73] *See* Columbia Global Freedom of Expression, case summary, *Kohls v. Bonta*; Eugene Volokh, *The Volokh Conspiracy* (blog) (2024–2025).

[74] *Kohls*, *supra* note 5; *see also Alvarez*, *supra* note 14.

protected humour and commentary. The court noted that deepfakes are a real problem, but not an issue for which speech should be prematurely banned.

The lesson is not that there is no regulation. It's that laws which broadly affect political speech, impose liability without proving that a speaker was actually aware of the statement being false and that the listener suffered actual harm are constitutionally suspect in a jurisdiction with robust protections for political speech, where platforms have no legal duty to engage in fact-checking. Disclosure obligations and very specific rules are more likely to survive.

6.2 *Raghav Chadha v. Ashok Kumar (John Doe)*: separating satire from malicious fabrication

In a decision of 1 July 2026 (CS(OS) 466/2026), the Delhi High Court (Justice Subramonium Prasad)[75] considered an application by a sitting Member of Parliament for an injunction against anonymous creators of AI-generated videos, voice clones and manipulated content. The plaintiff abandoned his claim for rights to his personality, and kept the case to defamation, according to a second report of the judgment.[76] The court reiterated that prior restraint is an exception in defamation cases, as in Bonnard v. Perryman and Bloomberg v. Zee,[77] and that a higher standard of criticism, including "satire," is expected of those in public life. It examined fifty-two items, determined that most of them were political criticism or satire, and sealed out six items determined vulgar, profane and malicious, in addition to revealing subscriber and IP information of the uploader. The court also reportedly highlighted that the use of AI to create content with intent to defame a person's dignity violates Article 21[78] and that, while a comprehensive regulation on AI is pending, the court will have to consider the specific facts of each case.

The case is instructive as it shows how the court does the case by case rather than blanket relief. Its approach, which distinguishes between criticism, satire or malicious fabrication, is much like the proportionality analysis employed in Europe[79] and is a blueprint that courts could follow to safeguard political expression while also intervening against the worst content. Note: This information is from a blog of a law firm, so please seek out the main text before relying on this alone.

6.3 TAKE IT DOWN: a fast remedy for a narrow category

The American federal law is the most obvious example of a fast removal remedy designed for a class of content that is of marginal expressive value.[80] The harm it addresses, non-consensual intimate imagery, is severe, immediate and largely irreversible, and the speech interest in such imagery is slight. Those attributes warrant quick resolution. The disadvantage of it is procedural: It doesn't have a counter notice system; It doesn't have penalties for abusive claims. The statute indicates that if the category is narrow, speed can be balanced with speech interests, but it must be balanced with safeguards against misuse added in. It would be much more difficult to defend the same design in the context of political or reputational material where falsity and harm are debatable.

[75] *Raghav Chadha*, *supra* note 6.

[76] *See* reports in *The Indian Lawyer* and *LiveLaw* on the Delhi High Court decision (July 2026).

[77] *Bonnard*, *supra* note 38; *Bloomberg*, *supra* note 37.

[78] The Constitution of India, 1950, art. 21; *Puttaswamy*, *supra* note 40.

[79] *Axel Springer*, *supra* note 53.

[80] TAKE IT DOWN Act, *supra* note 4, § 3; *see supra* Part 5.3.

7. Toward a Principled Framework for Balancing

The above analysis of these issues points to six principles that could form a framework for the design. They are provided as a compendium of the comparative material, and not as an expression of the law of any particular jurisdiction.

1. Calibrate the response to the category of harm. Measure the response to the type of harm. Deepfakes exist in many forms. The speech value of non-consensual intimate imagery, fraud and impersonation is low and the injury is serious, so there is a need for robust and quick remedies. Where the speech value is greater, and the question of falsity truly subject to debate, there is a case for slower, adjudicative procedures in cases of defamatory fabrications of public affairs. Of course, the greatest protection is afforded to satire and parody. There will be one rule for all synthetic media that will be wrong for most of them.

2. Prefer disclosure to prohibition. Don't put up with it, talk about it. The least intrusive ways in which the speaker can continue to speak and the audience can make judgements are in the form of labelling, provenance metadata, and context. They form the basis of the EU model and the Indian amendments of 2026.[81] They are not easily examined in a manner that is proven to be effective, and whether audiences will notice them and trust them are both empirical questions.

3. Require knowing falsity and concrete harm for liability. Demand knowledge of "falsity" and "concrete harm" to be imposed in order to decide the liability. The Sullivan, Falwell and Alvarez line of reasoning,[82] along with the European concepts of veracity and method,[83] all point to the same principle—in order to be held liable for a deceptive depiction, the creator must know or recklessly disregard the falsity of the depiction and cause concrete harm. Statutes which waive these elements open themselves up to invalidation or abuse.

4. Build procedural safeguards into every takedown mechanism. Establish procedural safeguards in all takedown processes. Notice and takedown should be accompanied by reasoned decisions, and provide time for the uploader to respond, penalties for knowingly false notices, transparency reporting and independent review. The principle of actual knowledge in India through a court or a government order and the lack of a provision for a counter-notice in the American statute exemplify two extremes.[84]

5. Prefer proportionate remedies, and treat prior restraint as exceptional. Do not seek disproportionate remedies and only consider prior restraint as an exception. Interim injunctions are more likely to disrupt speech than correction, labelling, de-amplification and damages.[85] If there is a need for interim relief, like in the case of intimate imagery or clear malicious fabrication, there should be a narrow and specific direction to grant the relief against a specific and identified piece of content rather than to grant interim orders of general nature against a category of content, as done by the Delhi High Court in Chadha.[86]

[81] AI Act, *supra* note 3, art. 50; IT Rules, *supra* note 2.

[82] *Sullivan*, *supra* note 45; *Hustler*, *supra* note 28; *Alvarez*, *supra* note 14.

[83] *Axel Springer*, *supra* note 53.

[84] *Shreya Singhal*, *supra* note 21; TAKE IT DOWN Act, *supra* note 4, § 3.

[85] *Bloomberg*, *supra* note 37; *Bonnard*, *supra* note 38.

[86] *Raghav Chadha*, *supra* note 6.

6. Treat public figures differently, but not infinitely. Have different standards for public figures but not unconditional. Harsher criticism and satire of public officials and candidates should be tolerated.[87] The tolerance does not last when it comes to the creation of fictional imagery of the intimate, or to a wilful forgery of the authenticity of evidence of fact, especially close to an election. The line between commentary and invention of conduct is drawn by the relevant line.

8. Unresolved Difficulties

The liar's dividend. There is no labelling that deals with the reverse issue of authentic content being called fake. The answer might be more about the infrastructure for authentic media, authenticating processes in newsrooms and in courts, and public media literacy than in regulation of speech.[88]

Technical limits. There are detection tools and watermarks which can be broken, weakened or avoided and an open-source model is not under the control of any provider. The EU's demand for robust and interoperable marking[89] shows that this is recognised, but uptake will be patchy and a labelling system can lead to a false sense of security for the unmarked content.

Over-compliance by platforms. The shorter the time frame and the bigger the punishment,[90] the more motivation to get it off the list. The greatest risks are applied to satire, political criticism and speech by marginalised groups, and it is difficult to identify because removals are not litigated.

Imbalance of remedies. So far, the Delhi High Court's orders relating to personality rights have been in favour of stars and public personalities who have the financial means to go to court.[91] There is often a dearth of speedy judicial solutions for ordinary victims – particularly women who have been targeted by sexualised deepfakes.[92] The issue with a democracy-oriented system is that it should focus on easy statutory complaints rather than depending on elite litigation.

Jurisdiction and anonymity. Creators and hosts can be located outside of the country,[93] which could lend weight to the platform side of things, but could also create a risk of differing rules around the country creating a tension for cross-border content.

Political capture. Governments can use any mechanism to classify content as false or synthetic for the purpose of opposing and criticizing it. Independent oversight, judicial review and transparency of requests for takedowns are vital protections,[94] especially when there is an election at stake.

9. Recommendations

Based on this analysis, the paper proposes the following recommendations for both legislators and courts.

[87] *Sullivan*, *supra* note 45; *R. Rajagopal*, *supra* note 41.
[88] Chesney & Citron, *supra* note 9.
[89] AI Act, *supra* note 3, art. 50(2).
[90] *See* IT Rules, *supra* note 2; TAKE IT DOWN Act, *supra* note 4, § 3.
[91] *Yuvraj Singh*, *supra* note 6; *Anil Kapoor*, *supra* note 10.
[92] Ajder *et al.*, *supra* note 1.
[93] IT Act, *supra* note 30, s. 75; DSA, *supra* note 55, art. 2.
[94] *Shreya Singhal*, *supra* note 21; DSA, *supra* note 55, arts. 15, 17.

1. Adopt a harm minimisation approach. Implement specific remedies for non-consensual intimate deepfakes and impersonation fraud and maintain the balancing of political and reputational claims within the traditional defamation/personality law.
2. Label all realistic synthetic media and some forms of clearly artistic and satirical media with disclosure and harmonize labelling requirements across jurisdictions to minimize disconnection of compliance.[95]
3. Narrow the definition of offences and civil wrongs, so that they involve a realistic depiction, knowledge or recklessness as to falsity and the actual harm, and expressly exclude parody, satire or commentary.
4. Implements protection measures in notice and takedown processes: Counter-notice, penalties for false notice, human review for disputed political content, public transparency reporting, and appeals process.[96]
6. Implement a rapid response system for election periods, limited content restrictions with independent adjudication and short, specific orders.[97]
7. Invest in infrastructure for the provision of real media that has provenances and authenticity, as supplementary measures to takedown regulation,[98] as well as in media-literacy programmes.
8. Expand access to justice by creating cheap and accessible complaint mechanisms, with legal aid being provided for victims and guidance to courts on keeping their "John Doe" orders as narrow as possible to avoid expansive coverage of lawful speech.
9. Commission an independent assessment of the impact of deepfake laws on harm reduction and speech suppression to ensure regulation is based on evidence, not expectations.

10. Conclusion

Deepfakes do not make a new argument between free speech and reputation, they are just exacerbating the one that already exists, adding the power of "evidence" with "digitally fast and scaled" distribution. This comparative material outlines three broad strategies: America's strong protection of speech,[99] where defamation and intimate-imagery laws have been narrowly drafted and survive the free-speech test while election-deepfake bans do not; Europe's approach to proportionality and disclosure,[100] in which laws and policies impose disclosure obligations and risk management measures on speech rather than prohibition; and India's approach to a rapidly expanding judicial protection of personality and dignity,[101] coupled with a new statutory regime of labelling and accelerated takedown, all in a constitution that explicitly gives authority to defamation laws to limit speech.[102]

None of the three is complete. The best compromise combines all three: disclosure as the minimum, limited liability for knowingly misrepresentative and harmful portrayals, quick but protected relief for the most serious harms and strong safeguards for satire and political criticism. The test of any rule is how it applies to the unpopular joke, the severe political caricature and the leaked but true recording, as well as

[95] AI Act, *supra* note 3, art. 50; IT Rules, *supra* note 2.
[96] DSA, *supra* note 55, arts. 16–21.
[97] *See Raghav Chadha*, *supra* note 6.
[98] AI Act, *supra* note 3, art. 50(2).
[99] *Sullivan*, *supra* note 45; *Hustler*, *supra* note 28; *Alvarez*, *supra* note 14; *Kohls*, *supra* note 5.
[100] *Axel Springer*, *supra* note 53; AI Act, *supra* note 3, art. 50; DSA, *supra* note 55, arts. 34–35.
[101] *D.M. Entertainment*, *supra* note 10; *Yuvraj Singh*, *supra* note 6; *Raghav Chadha*, *supra* note 6.
[102] The Constitution of India, 1950, art. 19(2); BNS, *supra* note 33, s. 356.

the malicious forgery. A system that shields individuals from fabricated truths while not giving officials and platforms a general authority over what is the truth will be good for both of the democratic good values that this paper has discussed.[103]

Note on sources and verification

None of the three is complete. The best compromise combines all three: disclosure as the minimum, limited liability for knowingly misrepresentative and harmful portrayals, quick but protected relief for the most serious harms and strong safeguards for satire and political criticism. The test of any rule is how it applies to the unpopular joke, the severe political caricature and the leaked but true recording, as well as the malicious forgery. A system that shields individuals from fabricated truths while not giving officials and platforms a general authority over what is the truth will be good for both of the democratic good values that this paper has discussed.

[103] *See supra* Parts 7 and 9.