BEYOND FAKE: REGULATING AI-GENERATED DEEPFAKES IN INDIA

Mrinal Gope, Christ University, Delhi

ABSTRACT

Deepfakes are synthetic images, audio and video recordings that convey a person's appearance and voice to say or do something that did not happen in reality. Their dangers go beyond celebrity impersonation: they encompass non-consensual intimate imagery, monetary fraud, election fraud, fabricated proof, reputational assaults and suppression of valid criticism via the risk of synthetic manipulation. Because of the overlapping nature of laws regarding privacy, defamation, obscenity, identity theft, cheating, copyright, and intermediary liability and constitutional speech, Indian law has traditionally responded to such behavior in a variety of ways. This paper explores that framework as it stood on 13 October 2026 and dwells on the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, which introduced an express regulatory framework for synthetically generated information (SGI). The amendments call for actions to prevent unlawful SGI, labelling and provenance systems for lawful SGI, stricter verification requirements for big social media intermediaries and shorter deadlines for the specified takedowns and complaints. The paper claims that the framework is a significant improvement, but that it still leaves uncertainty regarding the substantive liability of creators, tool providers and disseminators, does not adequately guarantee procedural justice, remedies for ordinary users, evidence preservation, or protection of parody, journalistic reporting and political speech. It also observes that the Digital Personal Data Protection Act, 2023 has a staggered commencement and majority of the substantive processing functions are yet to be implemented at the cut-off date. It is suggested that a rights-based and risk-based approach should be adopted: that is, a model that separates harmful deception from lawful synthetic expression, allocates obligations based on control and culpability, retains evidence, affords immediate remedy to victims and affords meaningful appeal in the event of an incorrect removal.

Keywords: Deepfakes, synthetically generated information, privacy, reputation, freedom of speech, intermediary liability, IT Rules, 2021, Digital Personal Data Protection Act, 2023, personality rights, and India.

1. Introduction

The concept of visual and audio evidence has been transformed by synthetic media. A video can be convincing and have a politician proclaim a policy that never took place; a relative could be a voice that is duplicated and asking for an immediate transfer of money; a person could be a face that is superimposed in an intimate scene without their knowledge. The technology isn't just cheaper for forgery forms, it makes them easier. It is scalable, replicable and can spread to millions before the person being deceived even knows. While a denial can't go as far as the original clip, search results, reposts and derivative copies can hold the injury long after the first copy is gone.

The legal issue is not if all deepfakes should be banned. Synthetic media can be used in dubbing, for the benefit of the disabled, for education, for entertainment, for satire, and for legitimate artistic experimentation. The question is how law would differentiate between deceptive/exploitative fabrication and an explicit statement of expression, and how it would balance the right of the person being portrayed as well as the public interest in information. The constitutional starting point is not: privacy versus speech, but: a structured effort to balance the Articles 19 (1) (a) and 19 (2) with Article 21, in a way that balances equality and procedural fairness (Article 14).

The law in force on 13 October 2026 is followed throughout this paper. It is important to highlight that the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, notified on 10 February 2026 (with effect from 20 February 2026), expressly address certain types of synthetically generated information. It would be wrong to claim that there are no rules in Indian law that have been specific to the concept of "deepfakes. The more specific question is whether or not the law is spread across various statutory regimes and whether the new intermediary obligations sit alongside unanswered questions of individual liability, scope and safeguards and enforcement.

The crux of the argument is that India is progressing from indirect governance to express synthetic media governance, which is yet to be completed. The 2026 amendments bring greater transparency and hasten the process, but a label cannot compensate for bad reputation, a three-hour takedown time frame will not ensure the preservation of evidence, and the broad prevention obligation can lead to over-removal. The law will not be about AI, it will be about harmful and deceptive uses of AI. It must also not allow for judicially carved out personality

rights to be the exclusive rights of famous people: a student, worker or private citizen may have a more hurtful and intimate offense committed to their personality through an intimate deepfake than by the unauthorised merchandising of a celebrity.[1]

2. Distinctive harms, meaning and forms of deepfakes

Media created or substantially manipulated using machine learning or other computational methods that have become commonplace and typically make it look like a real individual or real event is referred to as a 'deepfake'. The amended IT Rules employ the more generic term 'synthetically generated information'. Rule 2(1)(wa), now amended, will be directed at audio, visual or audio-visual information, including information that is artificially or algorithmically created, generated, modified or altered in a way that appears real, authentic or true, and is likely to be perceived as being indistinguishable from an actual person or event. Other functions that are generally considered good faith editing or accessibility are also excluded from the Rules where they do not materially alter the content.[2]

Deepfakes are hard to prove legally because of four reasons. First, the fabrication can not only copy an existing copyrighted product, but also one of its identities: a face, a voice, a mannerism or a social relationship can be copied even when no single source image is reproduced. Secondly, the harm is highly contextual. A labelled parody of a public figure and an undisclosed fabricated confession can have different legal implications, but can be applied using similar methods. Third, synthetic material can be created by one, shared by another and amplified by platforms that did not create it or be aware of the falsity. Fourth, detecting is probabilistic. The compression, editing, re-encoding processes can defeat forensic tools and legitimate content can be misidentified as synthetic.

Non-consensual intimate imagery is the most harmful type of personal injury that can occur. This can be created even if the person has not taken an intimate photograph or video. The victim may be mocked, threatened, extorted, fired, and experience conflicts within the family and constantly worry that the material will be released again. The formal separation of 'real' from 'fake' intimate imagery can't resolve the core privacy harms: that the 'image' is one of the person's identity and body that has been appropriated, usually without their permission, to

1 Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, Gazette of India, Extraordinary, pt. II, sec. 3(i) (10 February 2026).
2 Carlotta Vaccari & Andrew Chadwick, Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust, 6(1) SOCIAL MEDIA + SOCIETY (2020).

produce an apparently intimate representation of themselves. The rules that focus only on the truth of the image, or that the victim must prove that it is real, would therefore not capture the central harm of the technology.

Other types of deepfake abuse are not sexual in nature. A cloned voice can authorize a fraudulent transfer; a false video can be used to misrepresent a public official or a candidate and make inflammatory comments; a fake interview can be used to tarnish a professional's credibility; and fake documents can appear to be certificates, bank statements or official notices. The 2026 Rules clearly identify several of these situations as high-risk or prohibited types of SGI. The regulatory system should, however, be "harm-sensitive," meaning that a fabricated image of a fictional person is not the same as a person's forged confession, and a "news" story reconstructed from a synthetic cast is not the same as showing invented footage as authentic evidence.[3]

3. Research approach and constitutional framework

Research approach and constitutional framework is the third element.The third element is research approach and constitutional framework.

This involves doctrinal analysis of legislation, notified rules, constitution benchmarks and judicial rulings, including recent High Court rulings on the aspects of synthetic media and personality rights. It does not provide an empirical analysis of the number of deepfakes or takedowns. The focus is on the organisation of legal responsibilities, the constitutional boundaries, remedies to be available and doctrinal loopholes that have emerged and will remain after the amendments of 2026. Ministers' FAQs and advisories are given as explanations not as substitutes for the text in the Gazette.

Article 21 guarantees the right to life and right to personal liberty, and the Supreme Court has acknowledged privacy as a fundamental right based on dignity, autonomy and informational self-determination. The nine judge decision in Justice K.S. Puttaswamy (Retd.) v. Union of India has dispensed with the notion of privacy as a limited interest in physical privacy. A deepfake can intrude upon privacy in a number of ways: by stealing biometric or expressive traits, by creating a fake intimate scene, or by controlling how the person is portrayed to others. The constitutional interest is not restricted to the right to be secretive, but extends to agency,

3 INDIA CONST. arts. 14, 19(1)(a), 19(2) & 21.

dignity and resisting coercion to be exposed or misattributed.[4]

There is no complete privacy. Reputation is not a magic bullet for public debate, either. The Supreme Court in R. Rajagopal v. State of Tamil Nadu has upheld the right to privacy, and has also held that discussion of public records is protected and that such suppression by a public officer is not generally allowed, unless there is a compelling countervailing interest. In a case emphasising post-publication remedies and public interest, deepfakes are focused on the difference between a fabricated deceptive work, which falsely attributes conduct, and a work of commentary, inference, parody or reporting, which is recognisably expressive.[5]

Article 19(1)(a) safeguards speech and expression, except for the reasons specified in Article 19(2). The term 'false information' is not a stand-alone constitutional basis for curtailing speech. A deepfake can touch a variety of legal areas depending on its context and content, such as defamation, public order, decency or morality, incitement, sovereignty or other recognised legal grounds. But because a powerful complainant says that it's fake, it shouldn't be taken down just because it's a mere disagreement with a political claim, a satirical edit or an embarrassing but legal expression. A restriction must be legally authorized and be reasonable and proportionate.[6]

The proportionality concept formulated in Modern Dental College and subsequently used in other cases is whether a measure serves a legitimate objective, whether it is an appropriate means to that objective, whether it is necessary (otherwise a less restive equally effective will serve the same purpose), and whether it is proportionate (otherwise no benefit is obtained). This means there is a need to focus on the detection accuracy, the potential for false positives, the scope of the order, the identity of the decision-maker, and the need for a prompt appeal, as well as the possibility of a label or context note addressing the harm more specifically than a removal. It's more difficult to argue against blocking an entire account than to block a single illegal file, particularly if a lot of the account consists of legal speech that isn't related to the illegal file in question.[7]

The Supreme Court has also noted that both speech and public discussion are at the heart of democracy. In the case of "Ministry of Information and Broadcasting v. Cricket Association of

4 Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1.
5 R. Rajagopal v. State of Tamil Nadu, (1994) 6 SCC 632.
6 INDIA CONST. art. 19(1)(a), art. 19(2) & art. 21.
7 Modern Dental College & Research Centre v. State of Madhya Pradesh, (2016) 7 SCC 353.

Bengal", the Court has discussed the significance of the public's right to know. While expression cannot be taken away just because there is a threat that a hostile action might be taken, there must be a proximate and legally cognisable danger for a restriction to be justified, says S. Rangarajan v. P. Jagjivan Ram. Those principles, however, do not prevent the spread of harmful fabrications, but they help protect against a regime where 'deepfake' would be a catch-all tool of censorship.[8]

4. The Information Technology Act and the 2026 SGI framework

The 2026 SGI framework and the Information Technology Act.

For many harms related to the use of information technology, the main central legislation is the Information Technology Act, 2000. Section 66C relates to identity theft with reference to the use of another person's electronic signature, password or other unique identification characteristic for the purposes of fraud or dishonesty. Cheating by personation, via communication device or computer resource is covered by Section 66D. The penalties in section 66E cover certain instances of capture, publication or transmission of an image of a person's private part in situations that offend against their privacy. Sections 67 and 67A concern obscene material and material containing sexually explicit acts or conduct in electronic form and section 67B the specified sexually explicit or exploitative material involving children. The use of any provision is dependent on its elements, not all uses of AI will automatically be found to be offences.[9]

Section 69A enables the Central Government to make a direction to block public access to information on certain grounds set out in the S69A, on a prescribed basis and with safeguards in place. Conditions for exemption from liability of intermediaries under Section 79 depend on meeting the statutory and rule-based conditions. These provisions do various roles. Section 69A is a blocking power for the Government, Section 79 is a safe-harbour and the IT Rules have due diligence requirements for intermediaries. The two provisions cannot be interpreted as giving the power to declare the truth or falsity of all matters in dispute. The Supreme Court in Shreya Singhal v. Union of India interpreted the word 'actual knowledge' in Section 66A and also restricted the interpretation of Section 79 to a court order or a legally authorised

8 Secretary, Ministry of Information & Broadcasting v. Cricket Association of Bengal, (1995) 2 SCC 161; S. Rangarajan v. P. Jagjivan Ram, (1989) 2 SCC 574.
9 Information Technology Act, No. 21 of 2000, §§ 66C–66E, 67, 67A & 67B (India).

notification of the Government. When rules put on platforms fast and consequential duties, its insistence on precision is still relevant.[10]

The 2026 amendments introduce a specific layer for SGI in the IT Rules, 2021. They not only describe what SGI is, they also make clear that, in the context of the provisions concerning unlawful conduct which refer to 'information', they mean SGI as well. They also clarify that technically disabling access or removing access in accordance with the Rules is not a breach of the conditions of section 79(2) as such. The amendments thus seek to address an operational issue: platforms should not be dissuaded from taking steps they need to take to combat illegal synthetic content if they fear losing safe harbour protection for doing so.[11]

New Rule 3(3) imposes covered intermediaries a duty to take reasonable and appropriate technical measures to prevent users from creating, generating, modifying, publishing, transmitting, sharing or disseminating SGI that infringes on the law. The rule explicitly covers certain categories of material such as child sexual exploitative or abuse material, material of non-consensual intimate imagery, material relating to invasion of privacy of the body, false documents or electronic records, and deceptive false depictions of a person or real event likely to deceive users. These are meaningful changes as no longer a general complaint and takedown approach is expected, prevention is required at service levels.

The amended framework introduces a requirement for permitted SGI that there is a clearly and prominently displayed label or notice and, if technically practicable, a permanent metadata or a technical provenance mechanism with a unique identifier that allows the identification of the synthetic content and the computer resource employed to create or modify it. The disclosure can be prepended to the content for audio, or prominent in the display for visual and audio-visual material. The Rules also mandate that intermediaries do not allow tampering with or removal of the label and provenance markers. This is a transparency rule, rather than a truth rule: Provenance can indicate that a file was created or modified but it does not indicate a violation of the law or a deceptive intent on the part of the creator.

The responsibilities are more robust for important social media intermediaries (SSMIs). An

10 Information Technology Act, No. 21 of 2000, §§ 69A & 79 (India); Shreya Singhal v. Union of India, (2015) 5 SCC 1.

11 Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, as amended by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 (India).

SSMI shall require a user declaration that content is and is not SGI, implement reasonable and appropriate technical measures to verify the user declaration, and clearly mark content that it has determined to be synthetic. This combination is better than relying on self-declaration alone, but is fraught with practical problems of access to detection tools, treatment of remixes and lawful satire, and treatment of content of unknown origin. A declaration may be incorrect, a detector may be incorrect and the lack of metadata does not mean a video is not authentic.

The 2026 Rules significantly shorten the time for responding. An intermediary shall remove or disable access to specified information which is unlawful within 3 hours after receiving an order from a court or a reasoned intimation from an authorised officer or agency of the government in the prescribed manner. Grievances normally must be addressed within seven days and the designated expedited categories listed in Rule 3(2)(a)(i) within thirty-six hours; complaints about the categories listed in Rule 3(2)(b) (including certain intimates, nudity, sexual and morphed content) must be removed and/or disabled within two hours. Periodic information about obligations and consequences is also to be given to users at least every three months; for services supporting SGI, extra warnings. These accelerated timelines can be helpful to victims, but also make it more critical to have clear notices, evidence collection, and a narrow scope and accessible appeal.

Technical design is the key to the Rules' effectiveness in practice. A good label should be difficult to cut and should not be damaged by reposting (where possible) or be illegible on a small screen. No single proprietary system should be used to determine a plant's provenance. Platforms should document the content identifier, when it was uploaded, the complaint, what was done, and preservation measures, and limit access to intimate content. Automated measures should be in proportion to the functions of a platform and also its risk profile, and not a requirement for “perfect detection,” which in reality is impossible. Disputes with regard to the legal obligations should be settled based on the notified text and the parent Act and constitutional principles and not based on every illustration in an FAQ being an offence in itself.

5. Criminal liability: deception, sexual harm, forgery and defamation

Criminal liability: deception, sexual harm, forgery and defamation

From 1 July 2024, the Bharatiya Nyaya Sanhita, 2023 (BNS) will be the new general criminal

law of the country replacing the Indian Penal Code. It may apply in relation to a deep fake as a component of stalking, voyeurism, harassment, criminal intimidation, cheating, cheating by personation, forgery or defamation. Section 356 is about defamation, section 318 is about cheating, section 319 is about cheating by personation, section 336 is about forgery, section 340 is about documents/records being made to look genuine when they are not, and sections 77 to 79 are about voyeurism, stalking and conduct intended to offend the modesty of a woman. The exact charge should be based on the facts and statutory ingredients—it should not just be because the media is synthetic.[12]

Deceiving someone and using a fabricated audio message impersonating a bank employee or family member can assist with a cheating by personation case, if the other elements are met. Electronic or false certificates issued may create a forgery issue. The context, intention or knowledge of a false video used to incriminate a person may constitute defamation or another crime. A fabricated drawing of a real public figure which can be clearly identified as fictional and is not used to induce a person by deception or fraud should not be subjected to an offence aimed at deception or fraud. To prevent the effect of prosecution upon lawful synthetic expression, criminal law must have a culpability threshold and clear proof.

There's a difference between identity theft and impersonation. The criminalisation of imitation of a face or voice in section 66C of the IT Act is not limited to any particular type of face or voice imitation; section 66C is limited to imitation of a face or voice in the context of fraudulent or dishonest use of an electronic signature, password or other unique identification feature. Section 66D may be applicable to cheating through personation on a website, but has to be proven to include the elements of the statute. Likewise, section 66E is based on certain images of a private place and private situations, while section 67 and 67A are reliant on the nature of the transmitted information. The prosecution could proceed under an intimate deepfake that is covered by these provisions, the BNS or the SGI-specific Rules, but it is the prosecution's responsibility to state the path the deepfake takes, not to assume that any one provision applies to all deepfakes.

It is important to highlight that the Protection of Children from Sexual Offences Act, 2012 (POCSO) is relevant in the context of synthetic content involving a child. It sets up a dedicated child protection system with definitions and provisions regarding the use of children for

12 Bharatiya Nyaya Sanhita, No. 45 of 2023, §§ 77–79, 318, 319, 336, 340 & 356 (India).

pornographic purposes and the storage and possession of pornographic material. The 2026 Rules also include child sexual exploitative and abuse material in the list of high-risk content that intermediaries should try to avoid. The system should ensure that children are not required to pass around or submit the abusive material repeatedly as proof to ensure the system protects them. Complaints should be structured to allow for a copy to be saved as a hash to aid investigation and to limit additional exposure and access by platform personnel.[13]

The situation of criminal defamation is more complex. In Subramanian Swamy v Union of India, the Supreme Court affirmed the constitutionality of criminal defamation under the previous IPC and also held that there is a legitimate interest in reputation that is protected by Article 21. The current offence is in section 356 of the BNS. A deep fake that misrepresents what is said or done by a person might be one of the most effective types of reputational attack, as it may seem to provide "sensory proof." However the injury done by defamation ought to be weighed against risk of resorting to criminal process to punish criticism. Courts shall uphold the statutory exceptions and consider the "public good" and "good faith opinion" and other defences as relevant.

If the threat and demand constitute criminal intimidation or extortion, then the person whose voice or face is being used to collect money may also be a victim of criminal extortion or intimidation. Ordinary elements in the criminal law are not omitted from the digital layer. Investigators should capture original URLs, timestamps, account identifiers, metadata and surrounding communications; note if an upload was public or private; and, obtain the source file if possible. A screenshot might not be enough to determine the provenance, authorship, or entire circumstances of the publication, particularly if the file has been re-encoded or altered several times.[14]

6. Privacy, dignity, reputation and personality rights

Several different, but related interests are involved in deepfakes. There are distinct aspects of privacy that support autonomy, intimate life and control over disclosure; there are aspects of privacy that relate to reputation—how a person is perceived; there are aspects of privacy that relate to personality/publicity rights, which involve certain unauthorised uses of the person's identity and its commercial value; and there are aspects of privacy that relate to data-protection

13 Protection of Children from Sexual Offences Act, No. 32 of 2012, §§ 13–15 (India).
14 Bharatiya Nyaya Sanhita, No. 45 of 2023, §§ 308 & 351 (India).

law, which concerns processing of identifiable personal data. All of these interests are overlapping, but no one interest is a perfect alternative to another. Even if the deep fake is not widely viewed, it could cause injury to privacy and dignity.An intimate, humiliating deep fake could cause privacy/dignity injury, regardless of whether it was widely viewed. A fraudulent endorsement can most often result in commercial misappropriation and consumer deception. An engineered political communication can involve other issues of reputation, public order and democratic integrity, and engage public discourse.

In R. Rajagopal, the Supreme Court held that privacy is the right to be left alone and that the publication of a person's name or likeness without his consent can be an invasion of the right to privacy. It also granted breathing space for the publication of information gathered from public records and public-interest reporting. Subsequent teaching, particularly Puttaswamy, is rooted in 'autonomy' and 'dignity'. These principles underlie protection against fabricated intimate depictions and non-consensual biometric appropriation. Statutory, contractual, IP/common-law remedies are more likely to be available against private creators and platforms than against State action, however; and the availability and scope of a general privacy tort is less well developed than the constitutional language implies.[15]

In a series of Indian cases relating to commercial appropriation, courts have evolved the concept of personality and publicity rights. In D.M. Entertainment Pvt. Ltd. v. Baby Gift House, the court underlined the importance of an individual's right to control commercial use of his/her identity and cautioned against going too far in infringing upon caricature, lampooning and parody. The unauthorised use of a celebrity's image in adverts was deemed to be an action in the case of Titan Industries Ltd. v. Ramkumar Jewellers. Those decisions form a basis for relief in cases of fake endorsement and disembodied voices mimicking the user to lure in business or when unauthorised AI avatars are sold as products. But these decisions do not create a blanket prohibition on all use of the user's identity.[16]

In the case of Anil Kapoor v. Simply Life India, the Delhi High Court specifically spoke about the misuse of an actor's name, likeness, voice and persona through the use of AI, deepfakes, face-morphing etc. The case illustrates how the civil and personality-rights concepts already available can be used to quickly obtain injunctions for an identifiable person's misuse of his or

15 R. Rajagopal v. State of Tamil Nadu, (1994) 6 SCC 632.
16 D.M. Entertainment Pvt. Ltd. v. Baby Gift House, 2010 SCC OnLine Del 479.

her persona. The order should be read in context however: it was an order sought by a prominent actor on a particular factual record and was not a comprehensive statutory code for all deepfakes or a Supreme Court judgment that decides all the constitutional boundaries.[17]

In the case of Jaikishan Kakubhai Saraf alias Jackie Shroff v. The Delhi High Court granted interim protection to the actor granting him a prohibition on certain uses of his likeness by the specified unauthorised actors such as uses under AI chatbots and/or distorted videos. In Arijit Singh v Codible Ventures LLP, the Bombay HC dealt with the usage of AI voice models and how the voice of a singer can be cloned without his consent. Both issues demonstrate the need for a contemporary interpretation of a personality right that goes beyond photographs, as the focus of the appropriation might be a recognizable voice, vocal style or digital avatar. The orders also highlight the benefit of targeted relief, for example, the elimination of plaintiff-specific references that are not necessary to the elimination of an entire video that could be edited more narrowly.[18]

In other orders like interim protection granted by the Bombay High Court in the case of Karan Johar vs. Indian Pride Advisory and the order passed by the Delhi High Court in the case of Salman Khan vs. Ashok Kumar/John Doe, a trend can be seen that is beginning to address misuse of the identity through digital technologies. Interim orders may be useful in situations where the spread of synthetic media is rapid, but are not to be stated as final decisions on all matters in contention. The sound legal framework should lead to the proof of identifiability, unauthorised use, the injury or commercial appropriation and a fair account of the defence's expressive purpose. Courts must differentiate between a faux endorsement, fan art, and an AI service that platforms as a “commercial” service from the celebrity's voice and a non-commercial research demonstration.Courts must discern between a faux endorsement, fan art, and a non-commercial research demonstration from an AI service platformed as a “commercial” service of a celebrity's voice, and deliberate humiliation from criticism of a public show.[19]

The rights of the people show a bigger disparity. Many of the privacy cases concern celebrities who have a financially valuable personality and have the means to obtain an urgent injunction.

17 Anil Kapoor v. Simply Life India, 2023 SCC OnLine Del 6914.
18 Jaikishan Kakubhai Saraf alias Jackie Shroff v. The Peppy Store, 2024 SCC OnLine Del 4106.
19 Karan Johar v. Indian Pride Advisory Pvt. Ltd., Bombay High Court (interim order, 2024); Salman Khan v. Ashok Kumar, Delhi High Court (interim order, 2024).

But a student who has been the victim of an intimate deepfake or a teacher who has been misrepresented as making an offensive statement might need protection in an emergency. A predictable civil cause of action and an affordable emergency procedure, along with clear standards for non-commercial misuse, are important in privacy and criminal law, which have a broader application. Dignity injury, bodily privacy, and exposure to extortion may be serious even in the absence of a commercial persona and legislation should not equate these with harm.[20]

The rights of the person after death is also uncertain and cannot be taken to be equal to the rights of the living. In Krishna Kishore Singh v. Sarla A. Saraogi, the Delhi High Court denied a request by a deceased actor's family member to establish some family rights of privacy, publicity and personality. The judgment holds great promise for Parliament to make a clear and limited pronouncement on the protection – who is entitled to it, for how long, for what purpose etc. – rather than relying on uncertain and broadening notions of personality rights. [21]

7. Freedom of expression, satire and the risk of over-removal

This constitutional challenge is particularly sharp because of the persuasive power of a deepfake and the fact that there is a high likelihood of such a restriction being justified when the content is something that has not yet been established as an attack, but is rather satire or criticism. A parody in a democracy can make one's hypocrisy obvious, exaggeration is an acceptable form of political messaging, and reenactments and synthetic illustrations can serve to explain a story for the journalist. The context or disclosure, likely deception, injury and statutory grounds should be the basis of the legal question, not the mere fact that a machine created the media. A rule allowing it to be illegal to depict any synthetic representation of any identifiable public figure would have an undue chilling effect on speech.

Shreya Singhal stands in the centre as the case held that an offence framed in such a generic manner that subjective offence and uncertainty posed a danger to the ordinary language of the statute was unconstitutional. It also includes the concepts of intermediary takedown knowledge, as well as the legal precision it demanded, which are also applicable to the 2026 framework. The new SGI Rules are materially different, creating a category of realistic synthetic audio-visual material and obligations addressed to known high-risk categories and

20 Karan Johar v. Indian Pride Advisory Pvt. Ltd., Bombay High Court (India) (interim order, 2024).
21 Krishna Kishore Singh v. Sarla A. Saraogi, 2021 SCC OnLine Del 3424.

obligations towards labels for allowed audio-visual material. But this is not enough to make a previous amendment to the IT Rules redundant just because it was declared unconstitutional. However, the constitutional principles learnt from the case of Shreya Singhal remain: the use of vague terms, the delegation of wide powers, the lack of procedural protection and the disproportionate consequences give rise to chilling effects.[22]

The amendment that enabled the government fact-checking unit to flag information regarding the Central Government business as fake, false and misleading was struck down by the Bombay High Court in Kunal Kamra v. Union of India in 2023. The Court accepted that the expression was too vague and too broad and that the amendment infringed on Article 14, 19(1)(a) and 19(1)(g) etc., among its various objections. It was based on a different rule and this should not be interpreted as invalidating the 2026 SGI changes. Its wider implications are the importance of avoiding a loose net for the executive to please with a claim to ascertain truth and the power intermediary liability can wield when it comes to forcing platforms to remove content before a court can.[23]

Tata Sons Ltd. v. Greenpeace International serves as an example of the importance that Indian courts have given to public interest commentaries/expressive criticism. Ltd. v. Galactus Funware Technology Pvt. Ltd. warns against the abatement of caricature, lampooning and parody. The idea is that sometimes an AI label isn't sufficient to make something legal. A "labelled" deep fake can be defamatory, intrusive or fraudulent. The argument is that the lawfulness of a depiction should be judged by the correct legal test, which has context and proportionality, not by an assumption that synthetic depictions are illegitimate. [24]

Particular care needs to be taken with pre-publication injunctions. R. Rajagopal cautions against prior restraint and acknowledges remedies after the publication in appropriate situations. However, in viral deepfakes scenarios, there may be post-publication relief that's not enough when private material is being shared or fraud is underway. However, there is no justification for an unlimited order for future content unknown, for all accounts or for all uses of a public person's name. An order should: name the target material, state the basis of the order under the statute or common law, provide an opportunity for the affected party to seek

22 Shreya Singhal v. Union of India, (2015) 5 SCC 1.
23 Kunal Kamra v. Union of India, 2024 SCC OnLine Bom 5 (Bombay High Court).
24 Tata Sons Ltd. v. Greenpeace International, 2011 SCC OnLine Del 466; Galactus Funware Technology Pvt. Ltd. v. D.M. Entertainment Pvt. Ltd., 2017 SCC OnLine Del 12166.

modification, on the narrowest remedy that will stop ongoing damage. If a platform inadvertently blocks legal content, it should offer an easy-to-understand notice and restoration mechanism.[25]

The Supreme Court in the case of Cricket Association of Bengal and also in Kaushal Kishor vs State of Uttar Pradesh reiterates the plural constitutional approach. Misinformation should not be a catch-all phrase for state authorities or platforms and a person's right to reputation should not be trumpeted by simply stating their power or fame. Balancing the Constitution is not a slogan for picking a side, it's a reasoned account of which right is affected, the seriousness and likelihood of the harm, the legal basis for intervention, and why a less restrictive response will not be effective.[26]

8. Data protection, copyright and related civil remedies

The identification of a person by his or her face, voice, photograph or other identifiable feature is considered personal data where the processing of that data is digital. The Digital Personal Data Protection Act, 2023 sets out a framework for digital personal data, consent and uses that constitute a specified legitimate use, obligations of data fiduciaries and the rights of data principals. The date of the commencement of the Act, however, was staggered with notification G.S.R. 843(E) dated 13th November, 2025. While several duties are linked to the core provisions set out in sections 3-17, these are not yet in force as of 13 October 2026 and will only come into force 18 months after notification. It would therefore be incorrect to assume that all of the consent and fiduciary duties contained in the DPDP Act were already binding on the cut-off date. [27]

In the future, the DPDP framework can offer a helpful pathway in the case of unlawful collection of personal data, for training of a model, or for processing into identity material of synthetic persons. However, its current design does not explicitly establish a full-fledged deepfake-specific cause of action, and its future with freedom of expression, journalistic purposes, and training on publicly available data will require careful scrutiny. The Act also does not make all biometric features a separate ‘sensitive personal data’ category as had done

25 R. Rajagopal v. State of Tamil Nadu, (1994) 6 SCC 632.
26 Secretary, Ministry of Information & Broadcasting v. Cricket Association of Bengal, (1995) 2 SCC 161; Kaushal Kishor v. State of Uttar Pradesh, (2023) 4 SCC 1.
27 Digital Personal Data Protection Act, No. 22 of 2023, §§ 3–17 (India); Ministry of Electronics and Information Technology, Notification No. G.S.R. 843(E) (13 November 2025).

some of the earlier Indian rules. The primary legislative challenge is determining what constitutes processing of a person's image and/or voice, what kind of consent is required and/or what other legal basis is sufficient, how exceptions to this principle will be applied, and the nature of the remedies when the resulting information is synthetic (i.e., not a simple disclosure of the input). [28]

Depending on the nature of data, and the nature of action, the substantive DPDP provisions and the Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011 might continue to be applicable for qualifying body corporates and for sensitive personal data. These are highly restricted and do not constitute a general personality rights act. This is another reason why data protection is not sufficient to overcome the problem of deepfakes; there is a need for legislation that specifically tackles issues relating to synthetic representations, identity deception and non-consensual intimate content, rather than simply regarding every harm as a typical data breach.

The other one is an incomplete avenue by copyright law. The Copyright Act, 1957 provides protection for original works and sound recordings, and sections 38A and 38B provide specified rights and moral-rights protection for performers. Enhancing existing audio to make it more realistic, reproducing protected music work, or publishing a modified version of a recording could pose a problem of infringement or performer's rights on the facts. Particularly, the author has special rights concerning some distortions or mutilations of the work under Section 57. Copyright does not confer monopoly rights on anyone over his or her own face or voice per se. Copyright must be used in addition to, and not in lieu of, privacy and civil protection, as a deep fake may be based on someone's identity without reproducing or stealing a specific copyrighted photo or recording. [29]

A deep fake may be relevant if the common law action for passing off and the Trade Marks Act, 1999, apply in situations where the endorsee falsely suggests that a person has endorsed a product or service, particularly where the registered mark or commercial goodwill is being wrongfully used. The Consumer Protection Act, 2019 also deals with misleading advertisements and provides the Central Consumer Protection Authority with powers in relation to misleading advertisements in accordance with the provisions of the act. For fake

28 Digital Personal Data Protection Act, No. 22 of 2023 (India).
29 Copyright Act, No. 14 of 1957, §§ 38A, 38B & 57 (India).

celebrity endorsements and AI-generated investment promotions, these regimes can be helpful. However, someone might not have a simple trade-mark or consumer law claim against a person who is being impersonated in a non-commercial harassment campaign. There is still a need for a general and proportionate civil remedy. [30]

The Indecent Representation of Women (Prohibition) Act, 1986 may apply to some images of women in ads or publications, depending on the terms of the law and how the ads are disseminated. It did not have generative AI, face swapping or synthetic intimate imagery in mind. The POCSO Act is more applicable to the issue of child sexual exploitative material, the BNS covers specified sexual and reputational offences, and the IT Act covers electronic transmission. These provisions may overlap, but if they are fragmented they will leave practical uncertainty for police, prosecutors, victims and intermediaries. The government guidance should specify how to report intimate content, but not create a rule that any non-sexual artistic/political depiction by a publisher is presumptively illegal. [31]

Deepfakes pose more challenges for elections. Section 123(4) of the Representation of the People Act, 1951 deals with certain statements of fact about a candidate's personal character or conduct that are published with the necessary intention or consent and are likely to affect his/her prospects. It does not prohibit the creation of all synthetic political content, nor are its provisions necessarily applicable to other election and criminal laws depending on the facts. The main points of any electoral rule should be material deception, timing, reach and attribution, and safeguarding political satire, news reporting and commentary. There is nothing wrong with using a disclosure obligation to enhance transparency, but public agencies should not be given an unchecked power to decide political truth.[32]

9. Evidence, attribution and enforcement problems

The cases of deepfake typically end with no law being named, not because there is no law to name. A complainant can only have a screen recording, the URL may have been lost before the police could record it, a file may have been compressed or stripped of metadata, etc. and a model may have been accessed via foreign infrastructure. The prosecution needs to identify the creator versus the uploader versus the reposter versus a service provider and monitor each of

30 Trade Marks Act, No. 47 of 1999 (India); Consumer Protection Act, No. 35 of 2019 (India).
31 Indecent Representation of Women (Prohibition) Act, No. 60 of 1986 (India); Protection of Children from Sexual Offences Act, No. 32 of 2012 (India).
32 Representation of the People Act, No. 43 of 1951, § 123(4) (India).

these with varying knowledge and control. The appearance of someone's face in a clip does not imply that that was the first to be produced, and the spread of the virus does not indicate the first producer.

Bharatiya Sakshya Adhiniyam, 2023 regulates electronic and digital records and contains a special provision for the admissibility of electronic records, including section 63. A prescribed certificate, integrity protection and evidence from the person or system that creates a record may be necessary for a case. While digital forensic examination can be used to assist in determining inconsistencies or metadata and file history, there is no detector that is infallible. The party asserting the deepfake needs to establish the origin, chain of custody and authentication of the creation; the party asserting it was fabricated should not have to prove a negative to the party using the deepfake by relying on proprietary tools that are inaccessible to the party asserting fabrication. [33]

Procedural coordination is also crucial. The Bharatiya Nagarik Suraksha Sanhita, 2023 has the existing criminal-procedure framework, which includes section 173 of information regarding cognisable offences. There must be cooperation between local police officers, cyber units, and forensic laboratories, as well as with the grievance officers on platforms, in handling cybercrime complaints. A victim should not be forced to submit multiple grievances to various jurisdictions when content is stored, created and reposted via various services. A national standardised protocol may assist in the evidence preservation, referral of intimate-image cases, immediate requests for intermediaries, and notifications of the complainant when an intimate-image upload is removed from the internet. [34]

A "content removal order" and a successful prosecution are not synonymous. If there is no secure copy of the link, it will be lost if you quickly remove it. The updated Rules themselves consider the preservation of relevant information in appropriate circumstances when user accounts are suspended or terminated. Good practice should ensure that the removal of evidence from the public is separated from the retention of evidence for the purposes of confidentiality – stop distribution as soon as necessary, preserve content and metadata with controlled access, make sure there is a reason for the action, and only release the material to an authorised investigator or court. Copies should not be kept on file or circulated in the school

33 Bharatiya Sakshya Adhiniyam, No. 47 of 2023, § 63 (India).
34 Bharatiya Nagarik Suraksha Sanhita, No. 46 of 2023, § 173 (India).

without a good reason.

One of the challenges is anonymity and cross-border services. The model training data itself or the original file may be on a platform that is not located in a country of origin, a voice-generation tool could be hosted in one country, content could be uploaded from another country, and a re-uploader may not be easy to identify. Platform cooperation and mutual assistance may be slower than virus circulation. A policy with only a quick takedown timeframe would thus eliminate the obvious signs of the problem and not necessarily lessen re-offending. It should come with the ability to secure evidence in the preservation process, identify repeated uploads, report compliance in a transparent way, and have a defined process for gaining access to the subscriber data using lawful means.

The next step is to review the post-2026 framework and identify achievements and areas for improvement.

10. Assessing the post-2026 framework: gains and gaps

In four respects the 2026 amendments are a real improvement. They set the technical definition, they don't consider all AI generated content as deepfake content, they differentiate between "prohibited content", "unlawful content" and "Synthetic content" that can be shared if it is clearly labeled, they introduce some requirements on the provenance information and they introduce platform duties even before a complaint is filed. This is because the tighter responsibilities for SSMIs acknowledge that platforms with scale and technical capacity have different capacities than small forums. The rules also provide for a reduction in some response periods, recognizing that intimate-image abuse and deceptive impersonation can result in short-term and ongoing abuse. [35]

There are three gaps in the structure. First, the Rules mainly apply to intermediaries. The IT Act and other laws may impose liability on individual creators but there is no clear and comprehensive statutory definition of liability for creating, commissioning or disseminating a harmful deepfake. While the elements of existing offences and thresholds differ, civil personality-rights actions continue to develop through case law. Second, the technical requirements are difficult but not clearly articulated in terms of minimum auditing

35 Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, Gazette of India, Extraordinary, pt. II, sec. 3(i) (10 February 2026).

requirements, accuracy reporting, independent testing and the treatment of false positives. Third, victims are confronted with unequal procedures to compensation, to the attribution of their account and to a rehabilitation of their reputation and to protection against re-uploading in the future. The new framework enhances the front end response, but it does not provide a complete victim-remedy architecture.

It is also subject to a due-process challenge because of the extremely short time frame. Delay may be justified for intimate imagery, a two-hour period for which removing the specified material may be warranted for the potential to aggravate harm, but systems must establish that the complaint is substantial enough to not inadvertently disable lawful reporting or counter-speech. Three hours after legal requirement of actual knowledge could enable platforms to remove first and review later – especially if penalties and safe harbour consequences are ill defined. Not simply extending all deadlines is the correct answer. It is to establish a set of procedures for prompt response, and to have very clearly defined categories, exact notice requirements, evidence preservation, and a quick review process to restore legal material.

A second tension is that technical detection may lead to a substitute for legal judgment. A detector cannot alone conclude that a file is defamatory, satirical, fraudulent, in the public interest or likely to mislead in context. But many legitimate uses can be addressed adequately with a label. Removal should only be considered for material which is illegal or specifically excluded by the rule, and for material that is likely to have a significant impact on news reporting, political discussion and public interest speech. The distinction between detection of man-made origin and determination of legality must be clearly set out in statutory guidance and platform design.

A third gap is related to risk distribution. A strong voice of authority can often get fast injunctions, whereas a regular victim might have a hard time with the proof of harm, evidentiary burden and the legal cost. Platforms should have an easily accessible grievance mechanism in Indian languages and offer a trusted person/ a representative to complain – especially for children and for those who are the victims of sexualised fabrications. The State must provide financial support for trained cybercrime units and forensic assistance, and legal aid and support for victims where serious injury to privacy or dignity is caused by synthetic imagery. It needs to not require a victim to post more explicit material than they need to prove what has occurred.

11. Recommendations for a rights-centred legal model

Our recommendations for a rights-based approach to a legal model are as follows:

There is a need for Parliament to consider a focused statutory chapter on harmful synthetic impersonation or a well-crafted standalone legislation. It is not meant to ban the use of deepfakes technology itself. Rather, it should describe realistic synthetic impersonation, and harms like non-consensual intimate images, impersonation, fraudulent evidence, bogus endorsements and targeted reputation attacks that exceed a defined level of culpability. The law should differentiate between creation for private experimentation and publication, commercial use and deliberate or reckless dissemination which is likely to result in serious injury. It should also match definitions with the 2026 IT Rules for not having competing and contradictory tests.

Secondly, the law must create a civil remedy for the ordinary man that is victim centred, as well as for the celebrities. Relief should consist of removal, injunction, correction of disclosure, compensation if it is found to be a leak, and steps to prevent re-leaking of a confirmed illegal file. The law should define what consent means as a requirement for intimate or biometric representation, but then not construe consent as a blanket requirement for legitimate journalism, documentary purposes, public interest commentary, parody or art. The inquiry should include consideration of identifiability, expectation of privacy, context, intent, likely deception, actual harm and of a less restrictive alternative.

Third, it is imperative that Parliament and MeitY institute due-process protections along with the stringent timelines. The notice must describe the allegedly illegal URL or file, the type of harm, the provision involved and the complainant's grounds for complaint (including very restricted exceptions where disclosure of information would put the victim at risk). Platforms should maintain an action log, securely store the file, offer rationale to the uploader when appropriate, and offer a good appeal to a human reviewer. If a removal has been made by mistake it should be corrected promptly. Government intimations must be signed by an authorised officer, be reasoned, must be kept to a minimum of facts and figures, have an audit trail and be subject to independent or judicial review.

Fourth, labelling and standards should be interoperable. The technical marker should remain legible after reasonable transformations, where possible; and public interest synthetic media

should be capable of carrying contextual disclosures about the purpose, sources and limitations of the reconstruction, should be readily discernable for an ordinary viewer. Watermarks are not a sufficient protection: They can be cut off or faked. The government should foster open provenance of detection tools and independent evaluations thereof. The need for a label does not imply that there is harm in the content and the absence of a label should not be interpreted to mean that there is deliberate deception.

Fifthly, the criminal law should not introduce duplicate and ambiguous offences. The BNS, IT Act and special statutes should be mapped in official guidance and examples should be provided of the elements which are required for identity theft, personation, defamation, offences involving sexual content, offences involving stalking and voyeurism. The guidance for prosecution must distinguish malicious deception from a clearly disclosed synthetic expression and prosecution officers should note the specific representation which is said to be false, the evidence of knowledge or intent as required, the person affected and the legal consequence. This would assist police in the avoidance of a blanket term ‘deepfake' being used as an alternative to statutory analysis.

Sixth and finally, electoral protection must be tailored specifically for the special role and influence of synthetic political communication. The emphasis needs to be on material fabrication of candidate statements or events, attribution, the proximity to voting, the ability to influence voting choice and contextual correction. Each fast-track intervention should be based on a clear set of legal principles, be transparent and be subject to independent review. A government department should not be able to have an unchecked right to determine the truth of its own actions without any review, and the Kunal Kamra litigation exhibits the constitutional peril of that. Election safeguards should also apply to disclosure made by campaigns and synthetic endorsements, but should not limit normal satires and political criticism. 36

Seventh, public education and forensic capacity and transparent reporting should support legal implementation. Training must be provided to the police, prosecutors, judges and platform moderators and election officials about synthetic media and the limitations of evidence. Transparency reports must be published annually and include details of the number of SGI labels issued, the number of complaints received, the type of removal, processing times,

36 Representation of the People Act, No. 43 of 1951, § 123(4) (India).

appeals and reversals, and reupload and government requests, without revealing the victims. There should be mechanisms to audit the independent results under privacy protective arrangements. Model standards should be updated as techniques change, but substantive criminal offences should be left to the legislative framework and not be formulated through non-substantive technical advice.

Last but not least, coordination should be achieved between the implementation of the DPDP Act and deepfake regulation. Prior to the entry into force of the substantive provisions, the government needs to explain how voice and image data, the training of models, sensitive intimate content and resulting artificial outputs will be handled; what rights people will have; and what exemptions will be in place to protect journalism and artistic expression. The Rules and commencement notifications of 2025 must be read carefully—these are staged documents, so they are not a guide to the existence of all substantive data-protection duties at present. Fragmentation can be addressed through a coordinated approach, but must not be used to extend the scope of data protection to cover 'suppression of lawful expression' under Article 19.[37]

12. Conclusion

The deepfakes issue is problematic legally because they are altering the form of evidence used to determine identity, reputation and truth. A realistic synthetic recording can reveal the sexual humiliation, financial fraud, political manipulation or a false accusation scale and speed that existing remedies are not prepared to deal with. In response to this, Indian law has adopted constitutional rights, the IT Act, criminal offences, privacy and personality rights doctrine, and intellectual property rights and intermediary obligations. The February 2026 SGI amendments mark an important move towards clarifying definitions, duties for prevention, labelling and labelling obligations, obligations for significant social media intermediaries, and the significant shortening of response time for a limited list of harmful content.

37 Digital Personal Data Protection Act, No. 22 of 2023 (India); Ministry of Electronics and Information Technology, Notification No. G.S.R. 843(E) (13 November 2025).

REFERENCES:

1. Bharatiya Nyaya Sanhita, No. 45 of 2023 (India).

2. Bharatiya Sakshya Adhiniyam, No. 47 of 2023 (India).

3. Constitution of India (1950).

4. Digital Personal Data Protection Act, No. 22 of 2023 (India).

5. Information Technology Act, No. 21 of 2000 (India).

6. Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 (India), as amended.

7. Ministry of Electronics and Information Technology. (2025, October 22). Explanatory note: Proposed amendments to the IT Rules, 2021 in relation to synthetically generated information. Government of India.

8. Ministry of Electronics and Information Technology. (2025, October 22). Draft amendments to the IT Rules, 2021 relating to synthetically generated information. Government of India.

9. Press Information Bureau. (2025, April 4). Government of India taking measures to tackle deepfakes. Government of India.

10. Press Information Bureau. (2026, March 25). Government has issued multiple advisories emphasizing the observance of due diligence obligations under the IT Act and IT Rules. Government of India.

11. Press Information Bureau. (2026, February 15). India AI Governance Guidelines: Enabling safe and trusted AI innovation. Government of India.

12. K.S. Puttaswamy v. Union of India, (2017) 10 SCC 1 (India).

13. Shreya Singhal v. Union of India, (2015) 5 SCC 1 (India).

14. Anuradha Bhasin v. Union of India, (2020) 3 SCC 637 (India).

15. R. Rajagopal v. State of Tamil Nadu, (1994) 6 SCC 632 (India).

16. MySpace Inc. v. Super Cassettes Industries Ltd., 2016 SCC OnLine Del 6382 (India).

17. Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753–1820.

18. Citron, D. K. (2014). Hate crimes in cyberspace. Harvard University Press.

19. Citron, D. K., & Franks, M. A. (2014). Criminalizing revenge porn. Wake Forest Law

Review, 49, 345–391.

20. Paris, B., & Donovan, J. (2019). Deepfakes and cheap fakes: The manipulation of audio and visual evidence. Data & Society.

21. Westerlund, M. (2019). The emergence of deepfake technology: A review. Technology Innovation Management Review, 9(11), 39–52.

22. Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust. Social Media + Society, 6(1).

23. Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? Business Horizons, 63(2), 135–146.

24. Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). The state of deepfakes: Landscape, threats, and impact. Deeptrace.

25. Europol. (2022). Facing reality? Law enforcement and the challenge of deepfakes. Europol Innovation Lab.

26. European Parliament and Council. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.

27. European Parliament and Council. (2022). Regulation (EU) 2022/2065 on a single market for digital services (Digital Services Act). Official Journal of the European Union.

28. European Parliament and Council. (2016). Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union.

29. National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce.

30. National Institute of Standards and Technology. (2024). Reducing risks posed by synthetic content: An overview of technical approaches to digital content transparency (NIST AI 100-4). U.S. Department of Commerce.

31. Coalition for Content Provenance and Authenticity. (2024). C2PA technical specification. C2PA.

32. Partnership on AI. (2023). Responsible practices for synthetic media: A framework for collective action. Partnership on AI.

33. OECD. (2019). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449). Organisation for Economic Co-operation and Development.

34. UNESCO. (2021). Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization.

35. United Nations. (1948). Universal Declaration of Human Rights.

36. United Nations. (1966). International Covenant on Civil and Political Rights.

37. World Economic Forum. (2024). Global risks report 2024. World Economic Forum.

38. Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. Science, 359(6380), 1094–1096.

39. Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe.

40. Maras, M.-H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos. The International Journal of Evidence & Proof, 23(3), 255–262.

41. Rini, R. (2020). Deepfakes and the epistemic backstop. Philosophers' Imprint, 20(24), 1–16.

42. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. Minds and Machines, 28, 689–707.

43. Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

44. Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

45. Kashyap, S. (2025). The digital mirage: India's evolving legal battle against deepfake technology. SCRIPTed: A Journal of Law, Technology & Society, 22(2). https://doi.org/10.2218/scrip.22.2.2025.12004