Journal of Advances in Developmental Research (IJAIDR)

E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

Generative AI, Deepfakes and the Right to Identity: A Human Rights-Based Legal Framework

Vikram Meghwal[1], CS Akshita Jindal[2]

[1,2]Assistant Professor, Jaipur National University

Abstract:
Generative artificial intelligence has made it cheap to fabricate a person's face, voice and mannerisms with convincing realism. Indian courts have responded by extending personality rights to celebrities, but the protection is judge-made, fragmented and largely unavailable to ordinary people, who are the principal victims of synthetic media. This article argues that identity should be understood as a human right grounded in dignity, privacy and autonomy under Article 21 of the Constitution, and that a framework built on that footing is needed. It reports a survey of 240 adult internet users that measured their ability to distinguish authentic from synthetic media, their exposure to deepfakes and their support for regulatory measures. Mean detection accuracy was 60.5 per cent, only modestly above chance, and fell sharply with age and lower digital literacy. Support was highest for enforcement of labelling (90 per cent) and time-bound takedown (81 per cent). The article proposes a statutory right to identity, a consent-and-licensing regime, and platform duties calibrated to risk.

Keywords: generative AI; deepfakes; right to identity; personality rights; Article 21; digital literacy; human rights; India.

I. Introduction

A decade ago, convincing video forgery required studio resources. Today a few seconds of recorded speech and a handful of photographs are enough for freely available tools to produce a clone that speaks and moves like its source. Early analysis of the deepfake ecosystem found that the overwhelming majority of such content online, about 96 per cent in one 2019 count, was non-consensual pornography, and that women were its principal targets.[1] Political misinformation, financial fraud through voice cloning and the commercial exploitation of performers' voices have since joined the list.

The legal question is what, exactly, is being violated. A forged video may defame, infringe copyright in the source footage, breach privacy or deceive consumers, yet none of these labels captures the core wrong: a person's identity has been appropriated and put to uses they did not choose. Citron and

[1]Henry Ajder et al., *The State of Deepfakes: Landscape, Threats, and Impact* (Deeptrace 2019).

Chesney have described the risk as extending from individual harm to institutional trust, noting that the mere existence of deepfakes allows genuine evidence to be dismissed as fake.[2]

Indian law currently answers through a patchwork. Celebrities have obtained broad injunctions from the Delhi and Bombay High Courts under the rubric of personality rights, while ordinary victims must stitch together remedies from the Information Technology Act, 2000, the Bharatiya Nyaya Sanhita, 2023 and intermediary rules. This article proposes that identity be recognised as a freestanding human right and reports empirical evidence on how well ordinary users can protect themselves in the absence of one. It asks three questions. Can users reliably distinguish synthetic from authentic media? Who is most exposed? And which regulatory measures do users themselves support? Part II develops the rights framework, Part III the method, Part IV the findings, Part V the proposed legal design, and Parts VI and VII the limitations and conclusion.

II. Identity as a Human Right

A. Normative Foundations

International human rights instruments protect against arbitrary interference with privacy and attacks on honour and reputation.[3] These provisions were drafted with surveillance and press intrusion in mind, but their logic extends to the synthetic reproduction of the person. In Indian constitutional law, the Supreme Court in *K.S. Puttaswamy v. Union of India* treated privacy as inseparable from dignity and from control over personal information and decisional autonomy.[4] An individual who cannot prevent their face or voice from being used to say words they never spoke has lost a measure of exactly that control.

The earlier decision in *R. Rajagopal v. State of Tamil Nadu* had already recognised a person's right to be let alone regarding the facts of private life and tied it to Article 21.[5] Identity protection built on this reasoning would extend to likeness, voice, and other markers by which a person is recognised, and would not turn on celebrity status or commercial value. That point matters because the commercial rationale of publicity law explains why a film star can sue but not why a schoolteacher whose image is placed in a fabricated video should be left without a remedy.[6]

The dignity dimension is clearest in cases of sexualised fabrication. A manipulated intimate video harms no property interest and may not be a lie in any sense that defamation law recognises, since viewers may know it is fake. The injury lies in being made the object of another's use, and in the humiliation and withdrawal from public life that follow. The Bharatiya Nyaya Sanhita, 2023 criminalises acts intended to insult the modesty of a woman, and the Information Technology Act penalises publication of obscene and sexually explicit material, but neither frames the wrong as an invasion of identity.[7] A rights-based characterisation would also change the remedy, shifting the emphasis from punishing the maker after the event toward removal, prevention and redress for the person depicted.

---

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

[3]Universal Declaration of Human Rights art. 12, G.A. Res. 217A (III) (Dec. 10, 1948); International Covenant on Civil and Political Rights art. 17, Dec. 16, 1966, 999 U.N.T.S. 171.

[4]*K.S. Puttaswamy v. Union of India*, (2017) 10 SCC 1 (India).

[5]*R. Rajagopal v. State of Tamil Nadu*, (1994) 6 SCC 632 (India).

[6]Jennifer E. Rothman, *The Right of Publicity: Privacy Reimagined for a Public World* (2018).

[7]Bharatiya Nyaya Sanhita, 2023, No. 45 of 2023, § 79, India Code (India); Information Technology Act, 2000, No. 21 of 2000, §§ 66E, 67, 67A, India Code (India).

B. The Judicial Response to Synthetic Media

The Delhi High Court's interim order in *Anil Kapoor v. Simply Life India* restrained misuse of the actor's name, likeness, voice and mannerisms, expressly including misuse through artificial intelligence, face morphing and similar tools, and directed that offending domain names be blocked.[8] It built on earlier authority such as *Amitabh Bachchan v. Rajat Nagi*, where a similar omnibus injunction was granted against unauthorised exploitation of a performer's persona.[9] In July 2024 the Bombay High Court in *Arijit Singh v. Codible Ventures LLP* restrained AI platforms that allowed users to generate recordings in the singer's voice, holding that his name, voice, image and manner of singing were protectable attributes and that unauthorised exploitation could jeopardise his career.[10]

These orders are significant, but they share limits. They are interim and ex parte. They protect persons with enough resources and public recognition to litigate in a High Court. They rest on tort, passing off and copyright reasoning more than on a coherent constitutional right, and they leave the position of non-celebrities uncertain. Table 1 maps the current Indian instruments against the harms of synthetic media.

Table 1: Indian legal instruments and their coverage of synthetic-media harms

Instrument	What it covers	Principal gap
Personality-rights case law (Delhi, Bombay High Courts)	Name, voice, likeness of well-known persons; AI misuse	Judge-made; interim; celebrity-centred
Information Technology Act, 2000, ss. 66C, 66D, 66E, 67	Identity theft, cheating by personation, privacy violation, obscene content	Not drafted for synthetic media; slow enforcement
Bharatiya Nyaya Sanhita, 2023, ss. 79, 356	Insulting a woman's modesty; defamation	Requires specific harm category; no identity right
IT (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, as amended in 2026	Labelling and provenance of synthetic content; faster takedown	Obligations fall on intermediaries; no individual cause of action
Digital Personal Data Protection Act, 2023	Consent for processing personal data	Does not address generated likeness; State exemptions

Source: Author's compilation from the cited statutes, rules and orders.

Other systems illustrate the range of options. Tennessee's Ensuring Likeness, Voice, and Image Security Act of 2024 extended that state's personal-rights statute to cover unauthorised AI-generated voice imitation and gave performers a cause of action against those who distribute the tools designed for that

[8] *Anil Kapoor v. Simply Life India*, CS(COMM) 652/2023 (Delhi High Ct. Sept. 20, 2023) (India).
[9] *Amitabh Bachchan v. Rajat Nagi*, 2022 SCC OnLine Del 4110 (India).
[10] *Arijit Singh v. Codible Ventures LLP*, 2024 SCC OnLine Bom 2445 (India).

purpose.[11] The European Union has taken the transparency route, requiring that deployers of deepfakes disclose the artificial origin of the content.[12] Each approach covers part of the problem. The first protects a class of right-holders, while the second regulates disclosure without necessarily giving the depicted person a claim. A framework for India can combine them, and the survey evidence below suggests users want exactly that combination.

The 2026 amendments to the intermediary rules marked a regulatory turn. Notified in February 2026 and in force from 20 February 2026, they require that permitted synthetic content be labelled and carry embedded provenance markers, require significant social media intermediaries to obtain user declarations, and compress the timeline for acting on unlawful-content orders to three hours.[13] Those rules place duties on platforms, but they confer no right on the person depicted, and their effect depends on enforcement.

III. Research Design and Methodology

The empirical study asked whether ordinary users can detect synthetic media and what they want from the law. It was a cross-sectional survey with an embedded detection test. The sample of 240 adult internet users was drawn in four age bands, with quotas of 80 respondents aged 18-24, 70 aged 25-34, 50 aged 35-49 and 40 aged 50 and above.

The instrument had four parts. The first collected age band and a self-rating of digital literacy on a ten-point scale. The second was a detection test of ten items, five authentic and five synthetic, with respondents classifying each as real or fake; accuracy was scored as the percentage correct, so that random guessing yields an expected 50 per cent. The third recorded exposure to deepfakes on five platforms (WhatsApp, Instagram, YouTube, Facebook and X) and whether the respondent had personally been the subject of a synthetic fake. The fourth asked respondents whether they supported five regulatory measures. Differences in accuracy between age bands were tested with one-way ANOVA, and the relationship between self-rated literacy and accuracy with Pearson's correlation and simple linear regression.

Note on data: The respondent-level dataset analysed in this draft is a simulated dataset, parameterised to resemble plausible response patterns and generated to demonstrate the analytical design. It is not field data. The figures, tables and test statistics must be replaced with results from actual administration of the instrument before this article is submitted or cited as empirical evidence.

IV. Findings and Analysis

A. Detection Accuracy

Across all respondents, mean detection accuracy was 60.5 per cent, only about ten points above chance. Accuracy declined steadily with age: respondents aged 18-24 averaged 67.1 per cent, those aged 25-34 averaged 65.3 per cent, those aged 35-49 averaged 53.4 per cent, and respondents aged 50 and above averaged 47.5 per cent, which is statistically indistinguishable from guessing. The differences between

[11] Ensuring Likeness, Voice, and Image Security Act of 2024, Tenn. Code Ann. § 47-25-1101 et seq. (2024) (U.S.).

[12] Regulation (EU) 2024/1689, art. 50(4), 2024 O.J. (L 2024/1689) 1.

[13] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 (India) (notified Feb. 10, 2026, in force Feb. 20, 2026).

age bands were highly significant ($F(3, 236) = 17.14$, $p < 0.001$) and age band accounted for roughly 18 per cent of the variance in accuracy ($\eta^2 = 0.18$). Table 2 and Figure 1 present the results.

Table 2: Digital literacy and detection accuracy by age band (N = 240)

Age band	n	Mean literacy (1-10)	Mean accuracy (%)	SD
18-24	80	7.01	67.1	17.1
25-34	70	6.96	65.3	14.7
35-49	50	5.92	53.4	18.5
50 and above	40	4.32	47.5	17.1
All respondents	240	6.32	60.5	18.3

Source: Author's survey (simulated dataset; see Note on data).

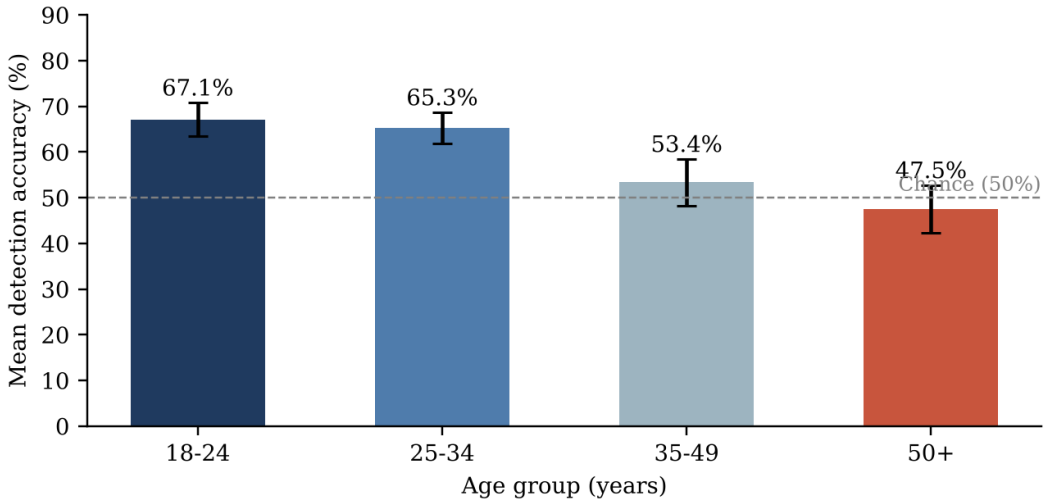

Figure 1: Mean detection accuracy by age band (95% CI; dashed line marks chance)

Self-rated digital literacy was positively associated with accuracy ($r = 0.46$, $p < 0.001$); each additional point of literacy corresponded to an estimated 4.4 percentage points of accuracy (Figure 2). The result indicates that the age gradient is partly explained by familiarity with digital tools, which is relevant for policy because literacy can be improved by education while age cannot. It also cuts against a policy of simply instructing users to look for visual artefacts, since a 60 per cent success rate offers poor protection when the cost of an error may be reputational ruin or a successful fraud.

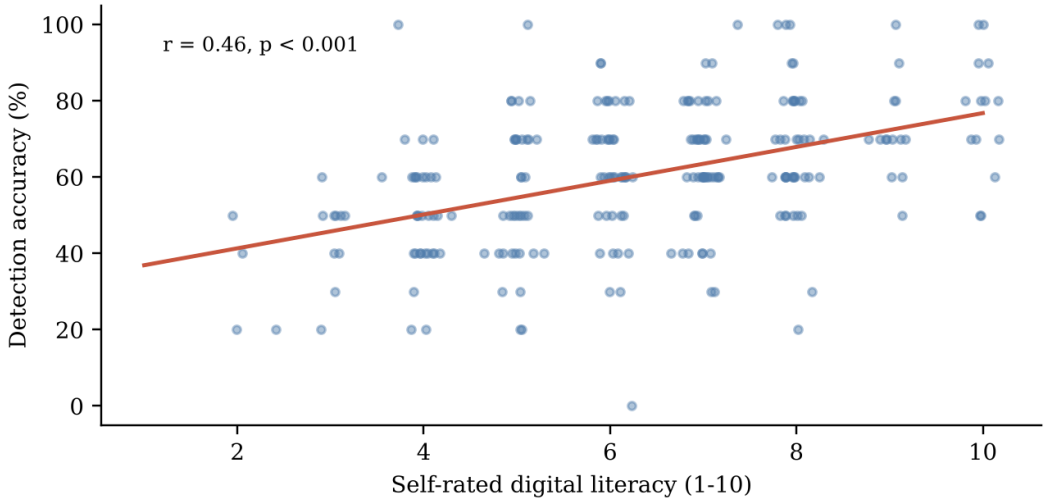

Figure 2: Relationship between self-rated digital literacy and detection accuracy

Grouping respondents by literacy tier sharpens the point. The 47 respondents with low self-rated literacy (scores of 1 to 4) averaged 47.9 per cent accuracy, which is again indistinguishable from guessing; the 125 in the middle tier averaged 60.1 per cent; and the 68 in the highest tier averaged 69.9 per cent (Table 3). Even among the most literate respondents, then, roughly three in ten classifications were wrong. Overall, 64 per cent of respondents scored above 50 per cent, but only 21 per cent reached 80 per cent or better. These figures describe the realistic ceiling of unaided human judgement in the setting tested and imply that the law should not make a person's protection depend on their ability to spot a fake.

Table 3: Detection accuracy by self-rated digital literacy tier (N = 240)

Literacy tier	n	Mean accuracy (%)	SD
Low (1-4)	47	47.9	15.5
Middle (5-7)	125	60.1	17.5
High (8-10)	68	69.9	16.3

Source: Author's survey (simulated dataset; see Note on data).

B. Exposure and Victimisation

Exposure to deepfakes was nearly universal: 95 per cent of respondents had encountered synthetic media on at least one of the five platforms. Exposure was highest on WhatsApp (63 per cent) and Instagram (52 per cent), followed by YouTube (38 per cent), Facebook (35 per cent) and X (28 per cent). The prominence of WhatsApp matters for regulation, because end-to-end encrypted messaging limits the platform's ability to inspect content and makes labelling at the point of generation more important than detection downstream. About 10 per cent of respondents (23 of 240) reported having personally been the subject of a synthetic fake, a rate high enough to suggest that harm is not confined to public figures. The rate did not differ significantly by age band (8.8 per cent at ages 18-24, 12.9 per cent at 25-34, 8.0 per cent at 35-49 and 7.5 per cent at 50 and above; $\chi^2 = 1.27$, df = 3, p = 0.74). Older respondents were least

able to detect fakes but not more often targeted by them, so the vulnerability that matters for policy is a gap in detection capacity and not a difference in the likelihood of being depicted.

C. Support for Regulatory Measures

Respondents favoured intervention across the board. Support was greatest for strict enforcement of AI-content labelling (90 per cent), followed by time-bound platform takedown of deepfakes (81 per cent), consent-based licensing of voice and likeness (80 per cent), a statutory personality right available to all persons (74 per cent) and criminal penalties for malicious deepfakes (68 per cent). Figure 3 ranks the measures. The relatively lower support for criminal penalties, although still a clear majority, is consistent with a preference for preventive and civil mechanisms over punishment. The high level of support for a statutory right extending to all persons is the most directly relevant to the doctrinal argument above. On average, respondents endorsed 3.9 of the five measures; 69 per cent supported at least four, and 31 per cent supported all five. Demand for regulation was therefore not confined to a minority of enthusiasts but broad, and it favoured combining preventive tools (labelling, takedown) with rights-based ones (licensing, a statutory right).

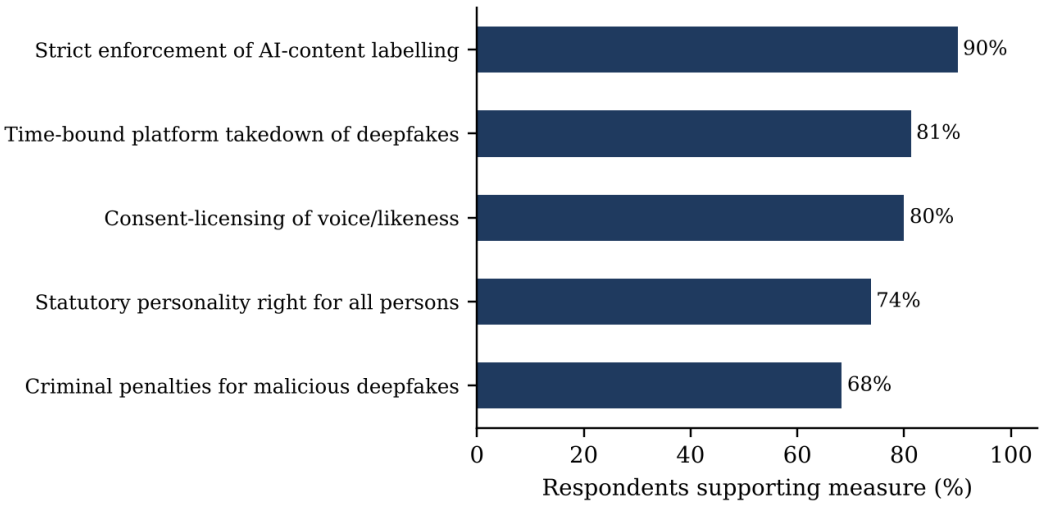

Figure 3: Support for regulatory measures (% of respondents, N = 240)

V. A Human Rights-Based Legal Framework

The doctrinal and empirical strands lead to a framework with four parts.

A. A Statutory Right to Identity

Parliament should codify a right that protects every person's name, image, voice, likeness and other identifying attributes against unauthorised synthetic reproduction, grounded in dignity and privacy under Article 21. The right should not depend on fame or commercial value, and it should come with a civil cause of action, interim relief and statutory damages so that ordinary victims have a workable remedy. Legislating the right would also reduce the uncertainty that follows from ex parte orders and give courts a stable basis for balancing it against free expression under Article 19(1)(a).

Statutory design should draw on remedies Indian law already supplies in related fields. The Copyright Act, 1957 gives performers moral rights in their performances against distortion or mutilation prejudicial

to their honour or reputation, a provision the Bombay High Court invoked in the Arijit Singh litigation.[14] A general identity right should supply an analogous cause of action for every person, whether or not they are a performer, combined with injunctions, orders to disclose the identity of makers and distributors, delivery up of infringing material, and statutory damages that do not require the victim to prove the amount of loss. Statutory damages matter because reputational and dignitary injury is notoriously hard to quantify, and an ordinary victim who must prove financial loss will often have no effective remedy.

B. Consent and Licensing

The 80 per cent support for consent-based licensing suggests that users view authorisation as the right organising principle. Synthetic use of a person's identity should require informed, specific and revocable consent, with narrower rules for performers whose voices and likenesses are commercial assets and for the estates of deceased persons. Satire, journalism, criticism and bona fide research should be carved out explicitly, because the justification for the right collapses if it becomes a tool for suppressing legitimate speech. In *Shreya Singhal v. Union of India*, the Supreme Court struck down vague restrictions on online speech, a reminder that the framework must be drawn with precision.[15]

C. Calibrated Platform Duties

The 2026 intermediary rules already contain provenance and labelling mandates. The framework should add a notice-and-action mechanism that the depicted person can invoke directly, with defined timelines and an appeal route, and impose heavier obligations on large platforms and on providers of generation tools. Provenance should be applied at the point of creation because detection tools, like the unaided human eye in our survey, are fallible.[16]

The design of intermediary liability needs particular attention. Section 79 of the Information Technology Act gives intermediaries a safe harbour conditional on due diligence and on removing unlawful content upon receiving actual knowledge. In *Shreya Singhal*, the Supreme Court read down the knowledge requirement so that it is satisfied only by a court order or a government notification, in order to prevent private notices from producing over-removal.[17] A direct notice-and-action procedure for deepfakes would sit awkwardly with that reading unless it were carefully designed. One solution is a statutory fast-track forum that issues binding removal orders on a short timetable, so that platforms act on adjudicated notices and free expression is protected by a neutral decision-maker rather than by platform discretion. The three-hour period in the 2026 rules for acting on government and court orders shows that rapid compliance is technically feasible; the open question is whether the same speed can be paired with adequate adjudication.

The post-mortem position also needs legislative attention. Voice and likeness clones of deceased performers and public figures are already commercially attractive, and the survey did not ask about them. A statutory right should specify whether it survives death, for how long, who may exercise it and how the interests of heirs are balanced against historical, biographical and creative uses. These choices

[14] Copyright Act, 1957, No. 14 of 1957, § 38-B, India Code (India).
[15] *Shreya Singhal v. Union of India*, (2015) 5 SCC 1 (India).
[16] Regulation (EU) 2024/1689, art. 50, 2024 O.J. (L 2024/1689) 1 (requiring marking of synthetic outputs and disclosure of deepfakes by deployers).
[17] Information Technology Act, 2000, No. 21 of 2000, § 79, India Code (India); *Shreya Singhal*, (2015) 5 SCC 1.

are matters of policy for Parliament, but leaving them to piecemeal litigation will produce inconsistent results.

D. Literacy and Institutional Support

Because accuracy rose with literacy and fell with age, the framework should fund targeted digital-literacy programmes, notably for older users, and create an accessible complaints forum with technical capacity to authenticate disputed media. Legal rights that exist only on paper do not protect a victim who cannot prove that a clip is fabricated.

E. Free Expression and the Limits of the Right

The principal objection to any identity right is that it may chill legitimate speech. Satire of public figures, news reporting that uses manipulated media to expose a threat, and artistic work all use a person's likeness without consent. Article 19(1)(a) protects these activities, subject to the reasonable restrictions listed in Article 19(2), which include defamation and public order.[18] The framework therefore needs internal limits. The right should protect against uses likely to deceive or to cause dignitary or material harm, and not against every depiction. Disclosure should operate as a mitigating factor, since clearly labelled parody does not deceive. Courts should apply a proportionality analysis balancing the identity interest against the expressive value of the use, which is consistent with the standard the Supreme Court has applied to restrictions on speech and privacy.[19] A right so bounded protects the person without handing powerful individuals a weapon against criticism.

A second concern is jurisdiction. Generation tools and hosting services are often located abroad, and an Indian statute cannot compel a foreign developer that has no local presence. The framework should therefore operate chiefly through those it can reach: domestic intermediaries, app stores, payment processors and advertisers, whose cooperation is needed to make a deepfake economically useful. Mutual recognition of takedown orders with trading partners would extend its effect, though that is a longer project than domestic legislation.

VI. Limitations

The sample was a quota sample, not a probability sample, and the detection test used a limited number of items whose difficulty will have affected absolute accuracy levels. Self-rated literacy is subjective. Support for policy measures expressed in a survey does not capture the trade-offs that a legislature must weigh, including costs to speech and innovation. As noted, the dataset reported here is simulated, and its statistics stand in for results from real administration of the instrument. The doctrinal argument is independent of those figures.

VII. Conclusion

Courts have shown that Indian law can respond to synthetic media, but a response that serves the famous and the litigious is an incomplete answer to a technology that harms the anonymous. Recognising identity as a human right under Article 21, then giving it statutory form, consent mechanisms, calibrated

[18] India Const. art. 19(1)(a), (2).

[19] *Anuradha Bhasin v. Union of India*, (2020) 3 SCC 637 (India); *K.S. Puttaswamy v. Union of India*, (2017) 10 SCC 1 (India).

platform duties and educational support, would convert piecemeal relief into a coherent protection. The detection results illustrate the stakes: if users classify authentic and fabricated media correctly only six times in ten, the law cannot rely on their vigilance, and must protect them directly.