

The Emerging Right Against Digital Manipulation: A Doctrinal Study of Human Rights in the Age of Generative AI

Vikram Meghwal¹, Princy Verma²

^{1,2}Assistant Professor, Jaipur National University

Abstract

Generative artificial intelligence can produce persuasive text, images, audio and video at scale, tailored to the individual and often indistinguishable from authentic material. This article contends that the harms of such systems are not adequately captured by existing rights to privacy, reputation or free expression, and that a distinct right against digital manipulation is emerging from international human rights law and from Indian constitutional principles. It defines the right through five components: identity integrity, informational authenticity, mental autonomy, algorithmic transparency and accessible remedies. The doctrinal argument is tested against two empirical exercises: a comparative coding of four jurisdictions on six legal indicators, and a survey of 96 legal academics, advocates and technology-policy professionals. Respondents endorsed all five components (means 3.86 to 4.47 on a five-point scale), with identity integrity ranking highest. The legal vehicle for the right divided opinion almost evenly between a standalone statute (40 per cent) and judicial recognition under Article 21 (38 per cent). The article proposes a layered recognition strategy.

Keywords: digital manipulation; generative AI; mental autonomy; freedom of thought; Article 21; human rights; algorithmic transparency.

I. Introduction

Human rights law has always had to adapt to the technologies that threaten the individual. The printing press produced protections for expression, the camera and the tabloid produced privacy, and mass data collection produced data protection. Generative artificial intelligence presents a different kind of challenge. It does not merely collect, publish or watch; it creates content designed to move particular people toward particular beliefs or actions, and it can do so using the person's own face, voice and behavioural data.

Existing rights capture part of this. A fabricated video of a politician may defame, a cloned voice may breach privacy, and a targeted advertisement may violate consumer law. Yet a common thread remains unnamed: the person's capacity to form beliefs and make choices free from covert, engineered distortion has been compromised. Susser, Roessler and Nissenbaum define online manipulation as the use of information technology to covertly influence another person's decision-making by targeting and exploiting their decision-making vulnerabilities.¹ Generative systems intensify that capacity because personalisation and content generation can now be combined in one pipeline.

This article advances three claims. First, there is a conceptual gap in human rights law that existing categories do not fill. Second, the gap can be bridged by recognising a right against digital manipulation

¹Daniel Susser, Beate Roessler & Helen Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World*, 4 Geo. L. Tech. Rev. 1 (2019).

with identifiable components, drawing on freedom of thought, privacy and dignity. Third, recognition in India is feasible by both judicial and legislative routes, and the legal community is divided about which should lead. The doctrinal analysis is supported by a comparative coding exercise and an expert survey. Part II defines the problem, Part III examines the normative sources, Part IV explains the method, Part V reports findings, Part VI proposes a recognition strategy, and Parts VII and VIII state limitations and conclude.

II. Defining Digital Manipulation

Not all influence is manipulation. Advertising, advocacy and political rhetoric attempt to change minds in ways the law has long tolerated. The distinguishing features of digital manipulation are covertness, personalisation and exploitation of vulnerability. A message is manipulative where the person cannot readily detect that they are being steered, where the steering is tailored to their psychological weak points, and where the mechanism bypasses their rational agency. Karen Yeung's concept of the *hypernudge* captures the architecture: big-data analytics continuously adjust the choice environment to the individual, producing influence that is dynamic, personalised and largely invisible.² Shoshana Zuboff's account of surveillance capitalism adds the commercial incentive.³

Generative AI adds three features. The first is synthesis: realistic falsification of evidence, including audio and video, which erodes the shared factual basis on which public deliberation depends. The second is scale and cost: personalised persuasive content can be produced for millions of recipients at negligible marginal cost. The third is interaction: conversational systems can build rapport and adapt in real time, which makes manipulation responsive in a way that static content cannot be. The combination is why the problem should be analysed as a distinct rights issue and not as a sum of defamation, fraud and privacy claims.

III. Normative Sources of the Right

A. International Human Rights Law

The strongest textual anchor is freedom of thought. Article 18(1) of the International Covenant on Civil and Political Rights protects freedom of thought, conscience and religion, and under the Covenant's structure the internal dimension of that freedom, the *forum internum*, admits no limitation, unlike the manifestation of belief.⁴ Article 19(2) protects the freedom to seek and receive information, which presupposes an information environment that is not systematically falsified, and Article 17 protects against arbitrary interference with privacy.⁵ Read together, these provisions support a claim that deliberately engineered distortion of belief formation touches rights the Covenant treats as fundamental, though the case law applying them to algorithmic influence is still sparse.

Soft-law sources point the same way. The Human Rights Committee has explained that the freedom to hold opinions without interference, and the freedom of thought, conscience and religion protected by Article 18(1), permit no limitation, and that the Covenant's protection of expression extends to the means by which ideas are received.⁶ A 2021 report of the Special Rapporteur on freedom of religion or belief examined the freedom of thought directly and warned that technologies capable of influencing thought raise questions the existing framework has yet to resolve.⁷ Academic work on cognitive liberty has argued for a similar

²Karen Yeung, *'Hypernudge': Big Data as a Mode of Regulation by Design*, 20 Info. Comm. & Soc'y 118 (2017).

³Shoshana Zuboff, *The Age of Surveillance Capitalism* (2019).

⁴International Covenant on Civil and Political Rights art. 18, Dec. 16, 1966, 999 U.N.T.S. 171; *see also* Universal Declaration of Human Rights art. 18, G.A. Res. 217A (III) (Dec. 10, 1948).

⁵International Covenant on Civil and Political Rights arts. 17, 19, 999 U.N.T.S. 171.

⁶U.N. Hum. Rts. Comm., General Comment No. 22, U.N. Doc. CCPR/C/21/Rev.1/Add.4 (July 30, 1993); U.N. Hum. Rts. Comm., General Comment No. 34, U.N. Doc. CCPR/C/GC/34 (Sept. 12, 2011).

⁷Special Rapporteur on Freedom of Religion or Belief, Interim Rep., U.N. Doc. A/76/380 (Oct. 5, 2021).

extension in the neurotechnology setting, and the reasoning transfers to systems that shape cognition through content and not through direct neural access.⁸ The Council of Europe's Framework Convention on Artificial Intelligence, opened for signature in 2024, requires parties to address risks to human dignity and individual autonomy, and to ensure accountability and remedies, which indicates movement toward treaty-level recognition of the underlying interests.⁹

B. Indian Constitutional Principles

Indian constitutional law offers several routes. In *Shreya Singhal v. Union of India*, the Supreme Court recognised that the open exchange of ideas is central to Article 19(1)(a), and in *Anuradha Bhasin v. Union of India*, it treated access to information and internet-based expression as protected and required proportionality for restrictions.¹⁰ *K.S. Puttaswamy v. Union of India* grounded privacy in dignity, autonomy and informational self-determination, and recognised that the right protects the sphere in which beliefs and choices are formed.¹¹ The decision in *Navtej Singh Johar v. Union of India* reaffirmed autonomy and dignity as constitutional values under Article 21.¹² Combining these strands, a right against manipulation could be framed as an aspect of the right to life and personal liberty, protecting the integrity of the processes by which a person forms views and decides.

Judicial practice has also begun to protect identity against synthetic misuse. Orders such as *Anil Kapoor v. Simply Life India* and *Arijit Singh v. Codible Ventures LLP* restrained AI-assisted exploitation of a performer's persona, demonstrating that courts can respond to generative AI under existing doctrine, although within the narrow frame of personality rights.¹³ The statutory side has begun to move as well. The 2026 amendments to the intermediary rules introduced labelling and provenance requirements for synthetic content and shorter takedown periods, but they impose duties on platforms without conferring a corresponding individual right.¹⁴

C. The Proposed Components

Drawing on these sources, the article defines the right against digital manipulation through five components. Identity integrity protects a person against synthetic falsification of their likeness, voice and conduct. Informational authenticity protects access to an information environment in which synthetic content is identifiable. Mental autonomy protects against covert, personalised influence that exploits psychological vulnerability. Algorithmic transparency entitles the person to know when and how automated systems are shaping what they see and are asked to do. Accessible remedies requires that enforcement be practicable for ordinary persons. The components are conceptually distinct but mutually reinforcing, and the survey described below measured support for each.

IV. Research Design and Methodology

The study used two empirical components alongside the doctrinal analysis. The first was a comparative coding of four jurisdictions, the European Union, China, the United States and India, against six indicators:

⁸Nita A. Farahany, *The Battle for Your Brain: Defending the Right to Think Freely in the Age of Neurotechnology* (2023).

⁹Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (opened for signature Sept. 5, 2024).

¹⁰*Shreya Singhal v. Union of India*, (2015) 5 SCC 1 (India); *Anuradha Bhasin v. Union of India*, (2020) 3 SCC 637 (India).

¹¹*K.S. Puttaswamy v. Union of India*, (2017) 10 SCC 1 (India).

¹²*Navtej Singh Johar v. Union of India*, (2018) 10 SCC 1 (India).

¹³*Anil Kapoor v. Simply Life India*, CS(COMM) 652/2023 (Delhi High Ct. Sept. 20, 2023) (India); *Arijit Singh v. Codible Ventures LLP*, 2024 SCC OnLine Bom 2445 (India).

¹⁴Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 (India) (notified Feb. 10, 2026, in force Feb. 20, 2026).

a prohibition on manipulative practices, a disclosure duty for synthetic content, a right protecting identity or likeness, algorithmic transparency obligations, liability for private actors, and the availability of an effective remedy. The author coded each cell from primary legal instruments on a three-point scale (0 = absent, 1 = partial or indirect, 2 = express and substantial). Because a single coder carried out the coding, the matrix should be read as indicative.

The second component was a cross-sectional survey of 96 respondents: 36 legal academics, 32 advocates and 28 technology-policy professionals. Respondents rated their agreement, on a five-point scale, that each of the five components should be recognised as part of a right against digital manipulation, and chose a preferred legal vehicle from four options: judicial recognition under Article 21, a standalone statute, amendment of existing data and information-technology legislation, or industry self-regulation. Group differences on each component were tested with Kruskal-Wallis tests, and vehicle preference across groups with a chi-square test. Significance was set at $p < 0.05$.

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 from the survey must be replaced with results from actual administration of the instrument before this article is submitted or cited as empirical evidence. The comparative coding in Section V.A reflects the author's reading of the instruments and is not simulated, but should be re-verified against current texts.

V. Findings and Analysis

A. Comparative Coding

Figure 1 reports the coding. The European Union scored highest (10 of a possible 12), reflecting the prohibition in the Artificial Intelligence Act on systems that deploy subliminal or purposefully manipulative techniques causing significant harm, its transparency duties for deepfakes and its requirements for high-risk systems.¹⁵ China scored 9, reflecting its rules on deep synthesis services, which require labelling and consent for use of biometric attributes, its interim measures for generative AI services, and the portrait and voice protections in its Civil Code.¹⁶ The United States and India each scored 4. In the United States, protection is fragmented across state right-of-publicity and deepfake statutes and a federal law aimed at non-consensual intimate imagery, with no general prohibition on manipulation or transparency mandate.¹⁷ India's score reflects judicially created identity protection, the new labelling mandate and general remedies, set against the absence of a manipulation prohibition or algorithmic transparency duty.

¹⁵Regulation (EU) 2024/1689, arts. 5(1)(a)-(b), 50, 2024 O.J. (L 2024/1689) 1.

¹⁶Provisions on the Administration of Deep Synthesis Internet Information Services (Cyberspace Admin. of China et al., Nov. 25, 2022, effective Jan. 10, 2023) (China); Interim Measures for the Administration of Generative Artificial Intelligence Services (Cyberspace Admin. of China et al., July 10, 2023, effective Aug. 15, 2023) (China).

¹⁷TAKE IT DOWN Act, Pub. L. No. 119-12 (2025).

European Union	2	2	1	2	2	1
China	1	2	2	1	2	1
United States	0	1	1	0	1	1
India	0	1	1	0	1	1
	Manipulation prohibition	Synthetic-content disclosure	Identity / likeness right	Algorithmic transparency	Private-actor liability	Effective remedy

Figure 1: Comparative coding of legal protection against digital manipulation
(0 = absent, 1 = partial, 2 = substantial)

Two observations follow. No jurisdiction scores the maximum on every indicator, and the weakest cell in every row is the remedy indicator, which scored 1 everywhere. The legislative advances in the European Union and China are concentrated in duties on providers and platforms and say less about the individual's enforceable claim. Only the European Union and China have anything approaching a prohibition on manipulative practices or a transparency duty, and the United States and India depend heavily on ad hoc doctrines.

B. Expert Support for the Five Components

Respondents endorsed all five components. Mean agreement ranged from 3.86 for mental autonomy to 4.47 for identity integrity. Table 1 gives the means, standard deviations and the share agreeing (scores of four or five). Identity integrity attracted the most support (93 per cent agreed), followed by algorithmic transparency (90 per cent), accessible remedies (85 per cent), informational authenticity (80 per cent) and mental autonomy (71 per cent).

Table 1: Expert agreement with proposed components of the right (N = 96)

Component	Mean	SD	Agree (%)	Kruskal-Wallis H	p
Identity integrity	4.47	0.63	92.7	9.57	0.008
Algorithmic transparency	4.27	0.64	89.6	2.69	0.261
Accessible remedies	4.12	0.64	85.4	5.23	0.073
Informational authenticity	4.12	0.71	80.2	7.02	0.030
Mental autonomy	3.86	0.66	70.8	2.91	0.233

Source: Author's survey (simulated dataset; see Note on data). H tests difference across the three respondent groups.

Mental autonomy, the least familiar component in legal vocabulary, drew the lowest support although still a clear majority. That ordering is consistent with a view that respondents find harms to identity easier to recognise than harms to belief formation, which are diffuse and harder to prove. It also matters for strategy:

the component that is most novel doctrinally is also the one on which professional opinion is least settled, and so is likely to require the most careful justification.

C. Differences Across Professional Groups

Differences across groups were significant for identity integrity ($H = 9.57$, $p = 0.008$) and informational authenticity ($H = 7.02$, $p = 0.030$), and not significant for mental autonomy, algorithmic transparency or accessible remedies at the 0.05 level. As Figure 2 shows, technology-policy professionals were more reserved than legal academics and advocates on identity integrity (means of 4.14 against 4.58 and 4.62) and on informational authenticity (3.82 against 4.28 and 4.22). A plausible explanation is that practitioners responsible for implementing duties are more alert to the cost and technical difficulty of verifying authenticity at scale. The convergence on transparency and remedies suggests that the groups share a core view of what a workable regime must include.

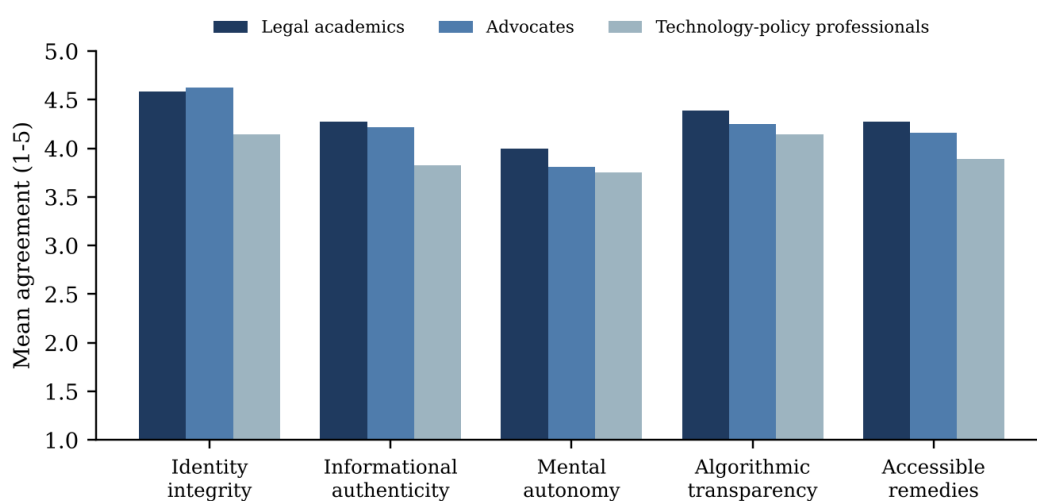


Figure 2: Mean agreement with each component, by respondent group

D. Preferred Legal Vehicle

Opinion on the route to recognition was divided. A standalone statute was the most common choice (40 per cent), narrowly ahead of judicial recognition under Article 21 (38 per cent); 16 per cent preferred amending the data-protection and information-technology laws and 7 per cent preferred self-regulation. Preference varied by group, with technology-policy professionals showing markedly more interest in amendment and self-regulation, although the association fell just short of conventional significance ($\chi^2 = 12.19$, $df = 6$, $p = 0.058$). Figure 3 presents the distribution. The near-equal split between statute and judicial recognition is the more important result, because it indicates that neither route commands a decisive majority and that a combined strategy is likely to be more acceptable than either alone.

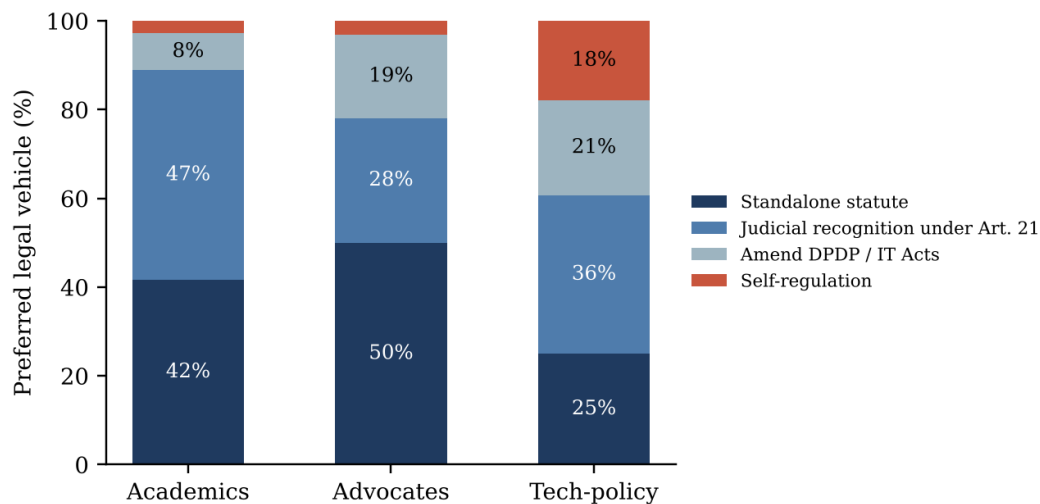


Figure 3: Preferred legal vehicle for recognising the right, by respondent group

VI. A Layered Recognition Strategy

The doctrinal and empirical results suggest a layered approach in which judicial and legislative action reinforce one another instead of competing.

A. Judicial Recognition as the Foundation

Constitutional courts can articulate the right incrementally. A claim framed under Article 21 would assert that deliberate, covert manipulation of a person's decision-making using automated systems infringes dignity and autonomy, and would be assessed through the proportionality framework set out in *Puttaswamy*. Judicial recognition has the advantage of speed and of building on the identity cases already decided. Its weakness, which the survey's lukewarm figures on mental autonomy reflect, is that courts work case by case and require a concrete factual record, which manipulation claims rarely supply.

B. Legislation to Supply Definitions and Duties

Legislation is better suited to define prohibited practices, allocate obligations and create enforcement bodies. A standalone statute, or a dedicated chapter in a future digital law, could prohibit systems designed to exploit psychological vulnerabilities in ways likely to cause significant harm, require disclosure of synthetic content and of the use of automated persuasion, and give individuals a direct cause of action. The European Union's prohibition on manipulative techniques offers a drafting reference, but its harm threshold and its enforcement through market surveillance authorities should be adapted to Indian institutions.

C. Enforcement Architecture

The comparative coding showed that remedies are the weakest element everywhere. A workable Indian regime should include an accessible complaints forum with technical capacity to authenticate disputed material, interim relief on a prompt timetable, a presumption in favour of the complainant where a provider cannot produce provenance records, and a duty on large providers to document risk assessments. Such mechanisms respond to the evidentiary asymmetry that makes manipulation claims hard to prove.

Existing instruments in related fields offer templates. The European Union's Digital Services Act prohibits online platforms from designing interfaces that deceive or manipulate users or materially distort their ability to make free and informed decisions, requires transparency about the main parameters of recommender systems, and obliges very large platforms to assess systemic risks to fundamental rights and civic

discourse.¹⁸ Indian data-protection law, by contrast, regulates the consent to processing and not the downstream use of personal data to shape behaviour. The Digital Personal Data Protection Act, 2023 could be amended to require that profiling for behavioural influence be disclosed and that consent for it be separately obtained, which would give the transparency component a statutory home without waiting for a new code.¹⁹ Combined with the labelling and provenance obligations in the 2026 intermediary rules, such amendments would assemble the beginning of a regime from existing legal materials.

D. Safeguards for Expression

A right against manipulation must not become a licence for censorship. It should expressly exclude persuasion that is transparent, satire and parody, journalism and artistic expression, and it should require proportionality in any restriction. *Shreya Singhal* and *Anuradha Bhasin* supply the constitutional discipline: vague, overbroad restrictions on online expression will not survive review.

E. Objections and Responses

Three objections are likely. The first is vagueness: manipulation is hard to define, and a vague right invites arbitrary enforcement. The response is to anchor the right in observable features, namely covert operation, personalisation to a known vulnerability and a likelihood of significant harm, and to confine the strongest remedies to those cases. The European approach of qualifying prohibited techniques by their purpose or effect and by the harm they cause is instructive on this point. The second is overlap: existing rights already protect privacy and expression, so a new right may be redundant. The reply is that the existing rights are triggered by particular harms, whereas the proposed right protects a precondition for exercising all of them, the integrity of belief formation and choice. That structural role resembles how the right to a fair hearing protects the conditions under which other legal rights can be vindicated.

The third objection is paternalism. Adults are entitled to be persuaded, and an anti-manipulation right could be used to treat all persuasion as suspect. The framework avoids this by targeting influence that is hidden and exploitative, not influence that is open and rational. A person who knows they are viewing an advertisement and weighs it is exercising autonomy. A person nudged by a system that infers and exploits their anxiety, without their knowledge, is not. The transparency component is intended to restore the ability to tell the difference, and so supports autonomy instead of substituting for it. The survey results fit this reading: transparency and remedies, which empower the individual, drew more support than mental autonomy, which asks the law to define the boundary of acceptable influence.

One further point concerns evidence. A person alleging manipulation will seldom have access to the logs, prompts, targeting parameters or training data that would prove it, and will often be unable to show that a particular piece of content was synthetic. Procedural rules should recognise that gap: courts or tribunals should be able to order disclosure of provenance records, draw adverse inferences when a provider cannot produce them, and appoint technical experts. Without such rules the right would be formally recognised yet practically unenforceable, which is the pattern the comparative coding found in the remedy indicator across all four jurisdictions.

VII. Limitations

The survey sample was purposive and small, and results cannot be generalised beyond the three professional groups. Agreement with a conceptual component does not show that respondents would support specific statutory text. The comparative matrix involved a single coder and a coarse three-point scale, and legal texts

¹⁸Regulation (EU) 2022/2065, arts. 25, 27, 34, 2022 O.J. (L 277) 1.

¹⁹Digital Personal Data Protection Act, 2023, No. 22 of 2023, India Code (India).

in this area are changing quickly. The survey dataset reported in this draft is simulated and its statistics must be replaced with field data. The doctrinal argument, which is the article's principal contribution, does not depend on those statistics.

VIII. Conclusion

The harms of generative AI are not limited to falsehood. They reach the conditions under which people form beliefs and make decisions. Existing rights respond to fragments of that harm, and the fragments do not add up to the whole. A right against digital manipulation, composed of identity integrity, informational authenticity, mental autonomy, algorithmic transparency and accessible remedies, gives a name and structure to the interest at stake. International freedom-of-thought doctrine and the Indian line of cases on dignity, privacy and expression supply its foundations, while the comparative evidence shows that remedies are the area most in need of design. Recognition will probably come in layers, with courts articulating the principle and Parliament supplying definitions, duties and institutions. What matters is that the interest is named before the technology outruns the law's capacity to protect it.