International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Consent in the Age of Correlation: Big Data Analytics and the Limits of India's Data Protection Framework

Vemuru Leela Krishna[1], **Astitwa Bhargava**[2]

[1]Research Scholar, Cyber Law And Information Security, National Law Institute University, Bhopal
[2]Assistant Professor, National Law Institute University, Bhopal

ABSTRACT

Big data analytics does not merely collect information; it manufactures it. Inferences drawn from aggregated datasets routinely reveal facts that no individual ever disclosed, and this inferential capacity sits uneasily with a data protection statute built around notice and consent. This paper examines whether the Digital Personal Data Protection Act, 2023 is equipped to govern big data processing in India. It begins with the constitutional baseline laid down in *Puttaswamy*, tests the consent-centric design of the 2023 Act against the realities of high-volume analytics, and then turns to two structural weaknesses: the statute's indifference to inferred and anonymised data, and the well-documented failure of anonymisation as a privacy safeguard. Comparative material from the GDPR and the American sectoral model is used selectively, not as an ideal to be copied but as a measure of where Indian law stands. The paper closes with concrete legislative and regulatory suggestions, including recognition of inferred data, a re-identification offence with teeth, and risk-based obligations for large-scale profiling.

Keywords: Big Data, Consent, DPDP Act 2023, Puttaswamy, Anonymisation, Profiling, Privacy.

I. INTRODUCTION

Somewhere in the last fifteen years, the unit of privacy harm quietly changed. It used to be the disclosure: a letter opened, a telephone tapped, a medical record leaked. Today the harm more often lies in the correlation, in what can be worked out about a person from a thousand innocuous data points that she never thought of as sensitive at all. Grocery purchases predict pregnancy. Mobility traces predict religion and caste. Battery consumption patterns, of all things, have been used to fingerprint devices. The individual pieces are trivial; the mosaic is not.[3]

Indian law has been slow to register this shift. The Information Technology Act, 2000 was drafted for e-commerce, not analytics.[4] Its data protection provisions, inserted almost as an afterthought in 2008,

[1]; Research Scholar, National Law Institute University, Bhopal. Assistant Professor, S.V.D. Siddhartha Law College, Kanuru, Vijayawada.

[2] Assistant Professor, National Law Institute University

[3]danah boyd & Kate Crawford, *Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon*, 15 INFO., COMM. & SOC'Y 662, 662–66 (2012).

[4]The Information Technology Act, 2000, No. 21, Acts of Parliament, 2000 (India). The long title speaks of "legal recognition for transactions carried out by means of electronic data interchange," which tells us what Parliament was worried about in 2000.

address discrete acts of wrongful disclosure and negligent handling of "sensitive personal data."[5] The Digital Personal Data Protection Act, 2023, India's first general data protection statute, is a considerable advance, but it is, at bottom, a consent statute.[6] Consent is the engine of lawfulness under section 6, and nearly everything else in the Act is plumbing around that engine. The question this paper asks is a simple one, though the answer is not: can a consent-centric statute govern a processing paradigm whose defining feature is that the most consequential uses of data are unknown, sometimes unknowable, at the moment consent is taken?

My argument proceeds in five steps. Part II describes what is analytically distinct about big data, because much confusion in the literature comes from treating it as ordinary data processing at larger scale. Part III sets out the constitutional baseline established in *Justice K.S. Puttaswamy v. Union of India*,[7] which any statutory scheme must satisfy. Part IV tests the DPDP Act against big data practice and finds it wanting in specific, identifiable ways. Part V draws comparative light from the GDPR and the American experience. Part VI confronts the anonymisation problem, which I regard as the single largest blind spot in the Indian statute. Part VII offers suggestions.

II. THE ANATOMY OF BIG DATA: WHY VOLUME CHANGES THE LEGAL QUESTION

Engineers define big data through the familiar Vs, volume, velocity, variety, later joined by veracity and value.[8] For a lawyer these definitions describe the wrong thing. They characterise the dataset; liability attaches to conduct. Four features of the conduct are what count.

The first is inference. What comes out of an analytics pipeline is not the input data but freshly manufactured information about the subject: a credit score, a churn probability, a health risk band. Tene and Polonetsky observed more than a decade ago that the centre of privacy gravity had moved from collection to use, and nothing since has proved them wrong.[9] The second is repurposing as a business model. Data is warehoused precisely because it may turn out to be worth something for a use nobody has imagined yet; Mayer-Schönberger and Cukier describe this as the option value locked inside stored data.[10] The third is relationality: processing my data changes what can be worked out about people who resemble me, so a decision framed as individual is in truth collective. The fourth is opacity, commercial (trade secret) and technical (models their own builders cannot fully explain).

Each feature attacks a different pillar of the classical scheme. Inference destroys the assumption that the subject knows what exists about her. Repurposing hollows out purpose limitation. Relationality exposes the individualism of consent as a category error. Opacity defeats transparency. A statute that engages none of the four has not regulated big data at all; it has regulated the world as it stood around 2005.

[5]The Information Technology (Amendment) Act, 2008, No. 10, Acts of Parliament, 2009 (India), inserting §§ 43A and 72A into the principal Act.

[6]The Digital Personal Data Protection Act, 2023, No. 22, Acts of Parliament, 2023 (India) [hereinafter DPDP Act].

[7]*Justice K.S. Puttaswamy v. Union of India*, (2017) 10 SCC 1 (India) [hereinafter *Puttaswamy I*].

[8]The formulation is usually traced to Doug Laney, *3D Data Management: Controlling Data Volume, Velocity and Variety*, META GROUP RESEARCH NOTE (Feb. 6, 2001); see also VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA 6–12 (2013).

[9]Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 251–56 (2013).

[10]MAYER-SCHÖNBERGER & CUKIER, *supra* note 7, at 98–106.

III. THE CONSTITUTIONAL BASELINE: PUTTASWAMY AND PROPORTIONALITY

Assessment of any Indian data statute begins, and to a surprising degree ends, with *Puttaswamy I*. Nine judges, speaking through six opinions but with one voice on the result, located privacy within the guarantee of life and personal liberty and read its protection across the whole of Part III, discarding the contrary authority of the 1950s and 1960s in the process.[11] Three strands of the judgment bear directly on this paper. The first is the recognition of informational privacy as a facet in its own right. The plurality was explicit that in the information age the threat comes as readily from private power as from the state, and it called on the Union to legislate a protective regime, which is the mandate the 2023 Act eventually answered.[12] The second is the proportionality framework: a measure trenching on privacy needs a legal basis, a legitimate objective, a rational and proportionate fit between end and means, and procedural safeguards against misuse.[13] The third, too little discussed, is the separate opinion of Kaul, J., which addresses profiling in terms and treats knowledge about a person as a form of power exercised over that person.[14]

Puttaswamy II then ran the proportionality test over an operating big data system. The Aadhaar architecture survived in the main, though the provision that had let private parties insist on Aadhaar authentication did not.[15] One may quarrel with the outcome, and the dissenting opinion remains, to my mind, the sharper treatment of aggregation risk, but the doctrinal position after the two decisions is not in doubt. Large-scale processing engages Article 21 whether the processor is public or private, and the governing statute must survive proportionality review, not merely a competence challenge. The DPDP Act therefore does not write on a blank slate; it writes on a constitutionalised one.

IV. THE DPDP ACT, 2023: A CONSENT-CENTRIC STATUTE IN AN INFERENCE-DRIVEN WORLD

A. Notice and Consent as the Load-Bearing Wall

Under the 2023 Act, processing is lawful only for a lawful purpose and only on one of two footings: the data principal's consent, or the narrow band of legitimate uses in section 7.[16] The consent provision itself stacks up the adjectives one expects of a modern statute, freeness, specificity, informedness, an unambiguous affirmative act, and confines the consent to data reasonably needed for the stated purpose.[17] As drafting, this would pass muster in Brussels. The trouble lies underneath the drafting. Solove's account of the self-management problem applies without modification: no one reads the notices, no one could read them all, and even a person who somehow evaluated each disclosure rationally would still miss the risk created by their combination.[18]

Big data sharpens every horn of that dilemma. The analytics firm values the data for uses that have not been conceived when the notice goes out, so the notice must be drafted either so widely that it says nothing or so narrowly that it is untrue. And because the injurious output is an inference rather than a disclosure,

[11] *Puttaswamy I*, *supra* note 6, overruling *M.P. Sharma v. Satish Chandra*, AIR 1954 SC 300 (India), and the majority in *Kharak Singh v. State of U.P.*, AIR 1963 SC 1295 (India).

[12] *Puttaswamy I*, *supra* note 6, at para. 185 (Chandrachud, J., plurality).

[13] *Id.* at para. 180; refined in *Justice K.S. Puttaswamy v. Union of India*, (2019) 1 SCC 1 (India) [hereinafter *Puttaswamy II*], at paras. 126–157 (Sikri, J.).

[14] *Puttaswamy I*, *supra* note 6, at paras. 585–588 (Kaul, J., concurring).

[15] *Puttaswamy II*, *supra* note 12, at paras. 355–367.

[16] DPDP Act, *supra* note 5, §§ 4, 6, 7.

[17] *Id.* § 6(1).

[18] Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1883–93 (2013).

the principal is being asked to authorise something she cannot inspect. In these conditions consent under section 6 functions as ritual: it moves legal risk from fiduciary to principal while moving no actual control in the opposite direction.

B. Purpose Limitation and the Problem of Secondary Use

The Act tethers processing to the specified purpose, and the specified purpose is whatever the fiduciary chose to write in its own notice.[19] A purpose limitation defined by the regulated party's own drafting carries its escape hatch within it: state the purposes loosely enough ("service improvement," "analytics and research") and formal compliance coexists comfortably with substantive defeat. The GDPR at least supplies machinery for the secondary-use question, directing attention to how closely the new purpose relates to the old, the setting in which the data was gathered, and what the individual could sensibly have anticipated.[20] The Indian Act has nothing comparable. The entire weight is thrown back on the notice, which is to say, onto the party with every commercial reason to draft it loosely.

C. The Exclusions: Inferred, Publicly Available and Anonymised Data

Three definitional choices do most of the damage. First, personal data is defined by reference to identifiability alone,[21] and derived or inferred data appears nowhere as a distinct category. The profile constructed about a person therefore attracts, at best, the same thin obligations as the raw material, and a respectable argument can be made that it escapes the consent framework altogether, since no one ever collected it from the principal. Second, the Act stands aside entirely where the principal has herself made the data public, a carve-out that reads like a legislative blessing on the scraping economy, including the scraping that feeds model training.[22] Third, and most consequentially, anonymised and non-personal data fall outside the Act altogether, and, unlike the 2019 Bill that preceded it, the enacted statute neither regulates how anonymisation is done nor punishes its reversal.[23] Part VI explains why that last omission is the one that should worry us most.

V. COMPARATIVE PERSPECTIVES: THE GDPR AND THE AMERICAN SECTORAL MODEL

It should be said plainly that the GDPR is not a big data statute either; Indian commentary sometimes writes as though Brussels solved the problem and Delhi merely failed to copy the answer. The Regulation's consent provisions suffer the identical self-management pathology.[24] What the Regulation does have, and the DPDP Act does not, are four working instruments: the secondary-use compatibility test already noticed; Article 22's shield against fully automated decisions with legal or similarly serious consequences, now read by the Court of Justice to reach the production of a credit score itself;[25] compulsory impact

[19]DPDP Act, *supra* note 5, §§ 5, 6(1); see also Digital Personal Data Protection Rules, 2025, r. 3 (itemised notice requirements).

[20]Regulation (EU) 2016/679 (General Data Protection Regulation), art. 6(4), 2016 O.J. (L 119) 1 [hereinafter GDPR].

[21]DPDP Act, *supra* note 5, § 2(t).

[22]*Id.* § 3(c)(ii). The GDPR contains no equivalent general exclusion; publicity is weighed within the art. 6(1)(f) balancing but does not expel the data from the Regulation. GDPR, *supra* note 19.

[23]The Personal Data Protection Bill, 2019, cl. 82 (India), had made unauthorised re-identification an offence; the 2023 Act carries no successor provision. The expert committee had warned of exactly this problem. COMMITTEE OF EXPERTS UNDER THE CHAIRMANSHIP OF JUSTICE B.N. SRIKRISHNA, A FREE AND FAIR DIGITAL ECONOMY 38–42 (2018).

[24]See Bert-Jaap Koops, *The Trouble with European Data Protection Law*, 4 INT'L DATA PRIVACY L. 250, 250–56 (2014).

[25]Case C-634/21, *OQ v. Land Hessen (SCHUFA Holding AG)*, ECLI:EU:C:2023:957 (Dec. 7, 2023); GDPR, *supra* note 19, art. 22.

assessment for high-risk processing under Article 35; and a threshold concept of personal data that asks whether identification is reasonably likely given the means realistically available, so that badly anonymised data stays inside the net.[26] The transfer jurisprudence, which struck down the Privacy Shield for want of adequate protection against state access, is a further reminder that big data governance and surveillance law can no longer be kept in separate rooms.[27]

The American picture instructs in a different register. There is still no omnibus federal statute; HIPAA, GLBA and FCRA share the field with an expanding layer of state legislation of which California's is the template.[28] Yet it is the U.S. Supreme Court that has engaged aggregation most candidly. *Riley* treated the sheer storage capacity of a smartphone as constitutionally transformative,[29] and *Carpenter* brought a week of cell-site location records within the Fourth Amendment despite the third-party doctrine, reasoning from what longitudinal location data reveals about the whole of a life.[30] Indian courts applying *Puttaswamy* hold all the doctrinal material needed to reason the same way about mosaic effects. What is missing is a statute that performs that reasoning against private fiduciaries.

VI. THE ANONYMISATION MYTH AND THE RE-IDENTIFICATION PROBLEM

The Act's indifference to anonymised data would be defensible if anonymisation were reliable. The record of the last twenty-five years says otherwise. Sweeney demonstrated at the turn of the century that three mundane attributes, postal code, sex and birth date, sufficed to single out the great majority of the American population, and made the point unforgettable by pulling the health record of a sitting Governor out of a supposedly de-identified hospital dataset.[31] Narayanan and Shmatikov later unmasked subscribers in the Netflix Prize corpus by matching it against public film ratings.[32] Ohm's synthesis drew the moral that has hung over the field ever since: a dataset can retain analytic utility or achieve genuine anonymity, but it cannot hold onto both at once.[33]

For India the consequence is blunt. A fiduciary holding a dataset squarely governed by the Act may pass it through a de-identification routine of unverified quality and step outside the statute completely, thereafter free to sell it, merge it, and, if profitable, reverse the process, because reversal attracts no penalty under the 2023 Act. In a jurisdiction operating the largest biometric identity programme in the world atop a fast-growing stack of health, payments and mobility platforms, this is not an academic gap.[34] The

[26]GDPR, *supra* note 19, recital 26; Case C-582/14, *Breyer v. Bundesrepublik Deutschland*, ECLI:EU:C:2016:779 (Oct. 19, 2016).

[27]Case C-311/18, *Data Prot. Comm'r v. Facebook Ireland Ltd. (Schrems II)*, ECLI:EU:C:2020:559 (July 16, 2020).

[28]California Consumer Privacy Act of 2018, CAL. CIV. CODE §§ 1798.100–1798.199.100 (West 2023), as amended in 2020; on the sectoral character of the American field, see Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583, 587–94 (2014).

[29]*Riley v. California*, 573 U.S. 373, 393–97 (2014).

[30]*Carpenter v. United States*, 138 S. Ct. 2206, 2217 (2018).

[31]Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* 2–3 (Carnegie Mellon Univ., Data Privacy Working Paper No. 3, 2000).

[32]Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets*, 2008 IEEE SYMP. ON SECURITY & PRIVACY 111, 111–14.

[33]Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1704 (2010).

[34]See *Puttaswamy II*, *supra* note 12, at paras. 260–274 (Chandrachud, J., dissenting), on aggregation and profiling risk in centralised identity architecture.

exemption operates as a self-certified exit door from the entire regime, and the incentives guarantee the door will be used.

VII. SUGGESTIONS

Five reforms follow, and I would rank them in this order. *First*, derived and inferred data should be brought expressly within the definition of personal data, whether by amendment or by authoritative interpretation from the Data Protection Board, so that the access, correction and erasure rights in sections 11 to 13 reach profiles and scores and not merely the collected inputs. *Second*, a re-identification offence on the model of clause 82 of the 2019 Bill should be restored, paired with a statutory yardstick for anonymisation built on the question whether identification remains reasonably likely with the means realistically available. *Third*, section 5 or the Rules should acquire a compatibility test for secondary use, so that purpose limitation ceases to be hostage to the fiduciary's own drafting. *Fourth*, the significant data fiduciary machinery in section 10, which already contemplates impact assessment, should be triggered by processing characteristics, scale, sensitivity, automated decision-making, rather than resting wholly on executive notification, and the assessments should be filed with the Board and open to its audit. *Fifth*, the publicly-available-data exclusion should be cut down to uses consonant with the context of the original publication, the contextual-integrity limit for which Nissenbaum has argued for years.[35]

VIII. CONCLUSION

None of this is nostalgia for the statutory vacuum that preceded 2023; the DPDP Act is a real achievement and deserves to be treated as one. But it is an achievement calibrated to the transaction, the discrete moment at which one person hands information to another, in an economy whose centre of gravity now sits in the correlation. The constitutional command after *Puttaswamy* is proportionate protection against the threat as it actually exists, not as it existed twenty years ago. Until the statute learns to see inferences, to distrust anonymisation, and to treat scale itself as a regulatory fact, consent will keep doing what it presently does: supplying lawful cover for the very asymmetries it was designed to correct. That is not data protection so much as its paperwork.

[35] HELEN NISSENBAUM, PRIVACY IN CONTEXT 127–57 (2010).