INTERNATIONAL JOURNAL OF LEGAL STUDIES AND SOCIAL SCIENCES [IJLSSS]

ISSN: 2584-1513 (Online)

Volume 4 | Issue 3 [May, 2026]| Page 176 – 188

THE CONSENT PARADOX: LARGE-SCALE ARTIFICIAL INTELLIGENCE TRAINING AND THE STRUCTURAL LIMITS OF INDIA'S DIGITAL PERSONAL DATA PROTECTION ACT, 2023

- Ananya Rai[1]

ABSTRACT

The Digital Personal Data Protection Act, 2023, fully operationalised by the Digital Personal Data Protection Rules, 2025, places individual consent at the centre of India's data-governance architecture. Yet the defining technology of the present decade — the large language model and other foundation systems trained on vast, indiscriminately gathered corpora — does not fit comfortably within a consent-first regime. This paper argues that the Act produces a structural paradox for artificial-intelligence training rather than a soluble drafting gap. The consent route is practically unavailable at the scale that model development demands, while the two doorways that might permit consent-free processing — the exclusion of publicly available personal data under Section 3(c)(ii) and the research exemption under Section 17(2)(b) — fail in opposite directions: read expansively, the former hollows out the very protections the statute promises; read strictly, the latter offers little shelter to commercial developers. The paper situates this tension within the right to informational privacy recognised in Justice K.S. Puttaswamy v. Union of India, contrasts India's deliberate omission of a "legitimate interest" basis with the European General Data Protection Regulation, and confronts the unresolved technical problem of erasing personal data already absorbed into trained model parameters. It concludes by proposing a calibrated framework — a purpose-bounded lawful basis for training, a narrowed exclusion, obligations graduated by scale, and rights of control reconceived for what the technology can deliver — that would reconcile innovation with the constitutional guarantee of privacy without collapsing into either prohibition or permissiveness.

[1] Third Year, B.A. LL.B. (Hons.), Institute of Law, Nirma University, Ahmedabad, Gujarat

Keywords: *Data protection; artificial intelligence; consent; publicly available data; informational privacy; Digital Personal Data Protection Act, 2023.*

I. INTRODUCTION

In the space of three years, two developments have reshaped the legal landscape of the Indian digital economy. The first is the enactment of the Digital Personal Data Protection Act, 2023, India's first comprehensive statute devoted to the protection of personal data, and its full operationalisation through the Digital Personal Data Protection Rules notified in November 2025.[2] The second is the arrival of generative artificial intelligence as a mass-market technology, built upon foundation models whose capabilities depend on training over corpora of staggering size and indiscriminate provenance. Each development is, on its own terms, a response to the same underlying phenomenon: the conversion of human life into machine-readable data. They were not, however, designed with one another in mind, and the friction between them has become one of the most consequential unresolved questions in Indian technology law.

The Act rests on a single organising idea — that the lawful processing of personal data should ordinarily flow from the free, informed and purpose-bound consent of the individual to whom the data relates. This is an attractive and intuitively sound premise. It treats the individual as the author of decisions about her own information and refuses to let commercial convenience override that authorship. But the premise was conceived for a world of identifiable transactions: a person signs up for a service, is shown a notice, and agrees to a defined use. The training of a large language model is not such a transaction. It involves the ingestion of billions of fragments of text, image and metadata, much of it scraped from the open web, in which the personal data of countless individuals is embedded without their knowledge and frequently without any prospect of obtaining their consent.

This paper argues that the encounter between the two produces a structural paradox rather than a mere gap in drafting. On the one hand, the consent pathway that the Act privileges is, for practical purposes, closed to the developer of a general-purpose model: it is impossible to seek specific and informed consent from the millions of unidentified persons whose data populates a training corpus. On the other hand, the only routes that would permit lawful processing without consent — the statutory exclusion of "publicly available" personal data and the narrow exemption for research —

[2]The Digital Personal Data Protection Act, 2023, No. 22 of 2023 (India). The Digital Personal Data Protection Rules, 2025, were notified and published in the Gazette of India on 14 November 2025, operationalising the Act through a phased commencement schedule.

are unsatisfactory at both extremes. Read generously, the publicly-available-data exclusion would license the wholesale appropriation of personal information so long as it once appeared online, draining the Act of meaningful content precisely where the risk to individuals is greatest. Read strictly, the research exemption affords no comfort to the commercial enterprises that build and deploy most consumer-facing models.

The argument proceeds in six steps. Part II reconstructs the architecture of lawful processing under the Act and explains why it leaves no general-purpose basis for model training. Part III examines the two doorways out of the consent requirement and shows why each is unstable. Part IV turns to the problem of erasure and the technical impossibility of cleanly removing personal data once it has been absorbed into model parameters. Part V draws a comparative contrast with the European Union's reliance on a "legitimate interest" basis. Part VI locates the debate within the constitutional guarantee of informational privacy articulated in Puttaswamy. Part VII offers a calibrated reform proposal. The paper concludes that India must choose deliberately among three futures — prohibition, permissiveness, or calibration — and that only the third honours both the promise of the statute and the realities of the technology.

II. THE ARCHITECTURE OF LAWFUL PROCESSING

To see why model training sits awkwardly within the Act, one must begin with how the statute structures the right to process personal data at all. The Act establishes that a person handling personal data — a "Data Fiduciary," in its vocabulary — may process it only on one of two footings: with the consent of the individual, or for a defined set of "legitimate uses."[3] There is no third, residual category. The Act deliberately abandons the open-textured language of "reasonable purposes" that featured in earlier drafts, and in doing so it removes the very elasticity on which data-hungry technologies tend to rely.

Consent is demanding. It must be free, specific, informed, unconditional and unambiguous, signified by a clear affirmative act, and confined to the particular purpose for which it is sought.[4] The Rules of 2025 reinforce this by requiring that the accompanying notice itemise, in plain language, the categories of data collected and the precise purpose of processing, and that it explain how consent may be withdrawn. Pre-ticked boxes, bundled permissions and the manipulative interface designs commonly

[3]DPDP Act, 2023, § 4 (recognising consent and certain "legitimate uses" as the only grounds for the lawful processing of digital personal data).

[4]Id. § 6; Digital Personal Data Protection Rules, 2025, r. 3 (form and manner of notice and consent).

described as "dark patterns" are expressly disallowed.[5] Consent is also revocable at will, and its withdrawal triggers downstream obligations to cease processing and erase data.

Apply these requirements to the training of a foundation model and the difficulty becomes plain. Consent must be specific to a purpose, but the defining feature of a general-purpose model is that its eventual uses are open-ended and unknowable at the moment of training. Consent must be informed by a notice delivered to the individual, but the individuals whose data is scraped from the web are neither identified nor reachable. Consent must be capable of withdrawal, yet — as Part IV explains — a model cannot meaningfully "forget" a particular person on demand. Each of the qualities that makes consent valuable as a protective device makes it unworkable as the legal foundation for indiscriminate, large-scale ingestion.

The alternative footing, the "legitimate uses," does not rescue the developer either. The provision lists specific, closed situations: data voluntarily provided by the individual for a purpose she has not objected to; processing by the State for the provision of benefits, services, licences or certificates; compliance with legal obligations or the orders of courts; responses to medical emergencies and threats to public health; and measures during disasters or breakdowns of public order.[6] Conspicuously absent is anything resembling a general commercial or innovation interest. Where the European regime offers controllers a flexible "legitimate interest" basis on which much AI development in Europe in fact proceeds, the Indian statute offers no equivalent. The omission was a conscious legislative choice, reflecting a preference for certainty and individual control over the discretionary balancing that "legitimate interest" entails.

The consequence is a regime that is, on its face, unusually strict towards the model developer. A purely domestic reading of these provisions suggests that training a model on the personal data of Indian residents is lawful only where each affected individual has consented to that specific use — a practical impossibility — or where one of the narrow legitimate uses happens to apply, which in the ordinary commercial case it does not. If the analysis stopped here, the conclusion would be that large-scale model training on identifiable Indian personal data is effectively prohibited. The analysis does not stop here, because the Act contains a prior question that displaces the entire consent inquiry: whether the data falls within the statute's scope at all. It is to that threshold question, and to the research exemption beside it, that the next Part turns.

---

[5]DPDP Act, 2023, § 6(1) (requiring consent that is free, specific, informed, unconditional and unambiguous, signified by a clear affirmative action); the Rules disallow consent procured through deceptive interface design or pre-ticked boxes.

[6]Id. § 7 (enumerating "legitimate uses," none of which encompasses a general commercial or innovation interest).

III. THE TWO DOORWAYS: PUBLICLY AVAILABLE DATA AND THE RESEARCH EXEMPTION

The Act defines the outer boundary of its application, and within that boundary lies the provision that does the most work in the AI debate. Section 3(c)(ii) declares that the Act does not apply to personal data that is "made or caused to be made publicly available" either by the individual to whom it relates or by another person who is under a legal obligation to publish it.[7] The drafting is spare, but its implications are vast. If a developer can characterise the contents of a web-scraped corpus as publicly available personal data, the entire edifice of consent, notice, purpose limitation and erasure simply falls away, because the data has been placed outside the statute altogether.

This is the first doorway, and it is seductively wide. A great deal of the personal data that populates training corpora — social-media posts, public profiles, comments, reviews, photographs, blog entries — was at some point placed online by someone. On a literal and generous reading, all of it would be publicly available and therefore beyond the Act's reach, leaving the developer free to ingest it without restriction. The attraction of this reading for industry is plain, and it has been pressed in submissions urging that the processing of publicly available data for model training be placed beyond doubt.[8]

Yet the doorway will not bear the weight that this reading puts on it, for at least three reasons. First, the exclusion is conditioned on the identity of the publisher. It applies only where the data principal herself made the data public, or where some other person did so under a legal obligation. A photograph that an individual posts to her own account may qualify; the same photograph re-uploaded by a stranger, scraped by an aggregator, or resurfacing on a third-party site years later does not, because the person who made it publicly available in that later instance was neither the data principal nor anyone under a legal duty to publish it. A developer harvesting the open web at scale has no realistic means of verifying, fragment by fragment, which of these conditions is satisfied — and the statutory text places that verification burden on anyone who would rely on the exclusion.

Second, the executive's own stated position is at odds with the expansive reading. The Government has indicated, in answer to a parliamentary question, that the scraping of publicly accessible personal data remains subject to the obligations of the information-technology framework and to the data-

[7]Id. § 3(c)(ii).

[8]See, e.g., the representations of industry associations to the Ministry of Electronics and Information Technology (2025) urging an express exemption for the processing of publicly available personal data for the training and fine-tuning of AI models.

protection statute's requirements of consent and transparency.[9] That position is difficult to reconcile with a literal reading under which publicly available data is simply outside the Act; but its very existence signals that the exclusion is not understood by the State itself as a blanket licence. The result is interpretive uncertainty at the precise point where developers most need clarity.

Third, and most fundamentally, an expansive reading would defeat the purpose of the statute. The harm that data-protection law guards against is not confined to the moment of disclosure; it arises from aggregation, inference and re-use. Personal data that is individually innocuous when posted can, once combined at scale and processed by a model, support intrusive inferences about identity, behaviour and belief that the individual never contemplated and could not have consented to. To treat the original act of publication as exhausting the individual's interest in her data — as the wide reading does — is to ignore the distinctive risks of large-scale automated processing. A doorway intended to spare the law from policing genuinely public information becomes, on that reading, an exit through which the most consequential processing escapes all scrutiny.

The second doorway is narrower and points in the opposite direction. The Act exempts processing necessary for research, archiving or statistical purposes from most of its substantive obligations, provided the processing conforms to standards prescribed under the Rules.[10] For genuinely academic or non-commercial work — the training of a model to study linguistic patterns, say, or to advance the public understanding of a disease — this exemption offers real shelter, and it sensibly recognises that research often cannot proceed on a strict consent basis. But it is shelter of limited extent. The exemption is tethered to a research purpose and conditioned on adherence to prescribed safeguards and traceable provenance; it does not extend to the commercial development and deployment of consumer-facing models, which is where the overwhelming majority of personal-data processing for AI in fact occurs. A start-up building a chatbot to sell to enterprises cannot plausibly recharacterise that activity as research within the meaning of the provision, and the conditioning standards are designed to keep the gate closed against such recharacterisation.

The two doorways therefore fail in complementary ways. The publicly-available-data exclusion is too wide if read literally — so wide that it would nullify the protections the statute exists to provide — and dangerously uncertain once that literal reading is resisted. The research exemption is appropriately

[9]Response of the Minister of State for Electronics and Information Technology in the Rajya Sabha (Aug. 2024), to the effect that the scraping of publicly available personal data remains subject to the Information Technology Act, 2000, the rules made thereunder, and the DPDP Act.

[10]DPDP Act, 2023, § 17(2)(b); the standards conditioning research, archiving and statistical processing are prescribed under the Rules.

principled but too narrow to accommodate the commercial activity that dominates the field. Between them lies a hollow middle: the ordinary case of a commercial developer training a general-purpose model on personal data drawn from the open web, for which the Act offers neither a workable consent route, nor a fitting legitimate use, nor a secure exemption. The paradox is now fully visible. The statute appears at once to prohibit the central activity of the AI economy and to permit it through an exclusion whose breadth no one is willing to defend.

IV. THE ERASURE PROBLEM AND THE MYTH OF REVERSIBLE TRAINING

Even if a developer surmounts the question of lawful basis, a further obligation of the Act exposes a deeper incompatibility between consent-centric data protection and the architecture of machine learning. The Act confers on the individual a right to the erasure of her personal data, and obliges the Data Fiduciary to delete that data once consent is withdrawn or the purpose for which it was collected is no longer being served, save where retention is required by law.[11] In the world of conventional databases this obligation is administratively burdensome but conceptually straightforward: a record is located and deleted. In the world of trained models it is something close to a category error.

A foundation model does not store its training data as discrete, retrievable records. It stores statistical regularities distilled from that data across billions of numerical parameters — the model's weights. A particular individual's data may have influenced the final values of those weights, but it is not lodged anywhere within the model as an excisable item. There is no row to find and strike out. To erase a person's contribution would, in the strict sense, require identifying and reversing the marginal influence that her data exerted on the training process — a problem the technical literature addresses under the heading of machine unlearning, and which remains, for models of any significant scale, computationally forbidding and frequently impossible without retraining the model from the beginning at enormous cost.[12]

This creates a mismatch the Act does not anticipate. The statute treats erasure as a routine entitlement, exercisable at the individual's option. The technology treats erasure as either impossible or ruinously

[11]Id. §§ 8(7), 12 (obligations and rights relating to the erasure of personal data on the withdrawal of consent or the fulfilment of the purpose of processing, subject to retention required by law).

[12]On the difficulty of removing the influence of specific training data from a trained model, see generally the technical literature on "machine unlearning," which indicates that exact unlearning in large models is generally infeasible without retraining from scratch.

expensive. A literal insistence on the right would, carried to its logical end, require a developer to retrain a model whenever a single data principal withdrew consent — an absurd and unworkable result no regulator is likely to demand. But the alternative, quietly tolerating non-compliance, drains the right of erasure of its meaning in precisely the context where personal data is processed most extensively. Commentary on the 2025 Rules has already flagged that developers will be expected to design systems capable of selectively removing data from training pipelines and logs, a demand that may be achievable for input datasets and operational records but not for the trained parameters themselves.[13]

The erasure problem also reframes the earlier question of consent. One reason the Act insists that consent be specific and revocable is to preserve the individual's ongoing control over her data. But control that cannot be exercised is illusory. If the withdrawal of consent cannot, as a practical matter, dislodge a person's data from a model already trained, then the consent obtained for training was never the meaningful, revocable consent the statute envisages; it was, at best, a one-way authorisation dressed in the language of choice. The difficulty is therefore not merely one of compliance logistics but goes to whether the consent model can perform its core protective function at all in the AI context. The lesson is cautionary for any reform: a solution that simply relocates training to a different lawful basis, without confronting the irreversibility of training, would address the entry point while leaving the exit sealed. Whatever framework governs the lawfulness of ingestion must be paired with realistic, technically literate obligations about what individuals can expect after their data has been absorbed.

V. A COMPARATIVE DETOUR: THE EUROPEAN "LEGITIMATE INTEREST" ALTERNATIVE

India's difficulty is sharpened, not diminished, by comparison with the European Union, whose General Data Protection Regulation served as a principal model for the Indian statute. The two regimes share a vocabulary of consent, purpose limitation, data minimisation and erasure. They diverge, however, on the single point that matters most for AI training: the availability of a flexible lawful basis that does not depend on consent.

Under the European Regulation, consent is only one of six lawful bases for processing. Alongside it sits the basis of legitimate interests — processing necessary for the legitimate interests pursued by the controller or a third party, except where those interests are overridden by the interests or fundamental

[13]Commentary on the 2025 Rules has observed that data fiduciaries will be expected to design systems capable of selectively removing personal data from training datasets and logs — a demand more readily satisfied for inputs and operational records than for trained model parameters.

rights of the individual.[14] This basis is not a free pass; it requires a structured balancing test in which the controller must weigh its own purpose against the reasonable expectations and rights of the data subject, and document that assessment. But it supplies exactly the elasticity that model training requires, and it is on this footing that much AI development in Europe in fact proceeds, subject to safeguards such as transparency and the right to object.[15]

The Indian Act, as Part II noted, contains no analogue. Its drafters chose certainty over flexibility, replacing the discretionary balancing of legitimate interest with a closed list of legitimate uses. The choice has a respectable rationale: legitimate-interest balancing is notoriously indeterminate, places the assessment in the hands of the very party that benefits from a permissive outcome, and has generated extensive litigation in Europe. But the cost of that certainty is now apparent. By foreclosing the balancing route, the Indian statute leaves the model developer without the very mechanism on which European developers rely, and pushes the entire weight of the AI question onto the publicly-available-data exclusion — a provision wholly unsuited to bearing it.

The European experience is instructive in a second respect. Even with a legitimate-interest basis available, European regulators have not treated AI training as unproblematic. A national supervisory authority temporarily suspended a prominent generative-AI service in 2023 over concerns about the lawful basis for processing training data and the absence of adequate information to users, and the episode prompted a broader regulatory reckoning across the continent.[16] The lesson is that a flexible lawful basis is necessary but not sufficient: it opens the door to lawful training while still requiring transparency, the ability to object, and attention to the special risks of large-scale processing. The Union has since layered onto its data-protection regime a dedicated Artificial Intelligence Regulation that imposes transparency and disclosure obligations on general-purpose and generative systems, including duties to label synthetic content and to publish summaries of training data — obligations directed not at the lawfulness of ingestion but at the accountability of the resulting system.[17]

For India, two conclusions follow. First, the absence of a legitimate-interest basis is not a minor drafting detail but the structural feature that creates the paradox; any serious reform must confront whether to introduce a calibrated equivalent. Second, the European trajectory shows that the lawful-

[14]Regulation (EU) 2016/679 (General Data Protection Regulation), art. 6(1)(f), 2016 O.J. (L 119) 1.

[15]Id. art. 21 (right to object to processing founded on legitimate interests); art. 17 (right to erasure).

[16]Italian Data Protection Authority (Garante per la protezione dei dati personali), measure of 30 March 2023, temporarily restricting the processing of Italian users' personal data by a generative-AI service pending compliance.

[17]Regulation (EU) 2024/1689 (Artificial Intelligence Act), arts. 50, 53 (transparency duties for generative and general-purpose AI, including the labelling of synthetic content and the publication of summaries of training data).

basis question and the systemic-accountability question are distinct and both require answers. India has begun to address the second through subordinate instruments — including amendments to the intermediary rules in 2026 that regulate synthetic and deepfake content,[18] and non-binding governance guidelines issued in late 2025 that favour a light-touch, principles-based approach[19] — but it has not yet resolved the first. A regime that regulates the outputs of AI while leaving the lawfulness of its inputs in a state of paradox addresses the symptom and neglects the cause. It would be a mistake, however, to treat European law as a template to be copied wholesale; the legitimate-interest basis has imported its own uncertainties, and the Indian preference for individual control reflects a defensible constitutional sensibility. The comparative lesson is not that India should adopt the European solution, but that it cannot avoid the European question.

VI. THE CONSTITUTIONAL FRAME: PUTTASWAMY AND INFORMATIONAL SELF-DETERMINATION

The statutory paradox does not arise in a constitutional vacuum. The Supreme Court's decision in Justice K.S. Puttaswamy v. Union of India recognised the right to privacy as intrinsic to the right to life and personal liberty under Article 21 of the Constitution, and it expressly located informational privacy within that guarantee.[20] The Court held that an individual's interest in controlling the dissemination and use of information about herself is a constitutionally protected dimension of personal autonomy, and it established that any intrusion on that interest must satisfy a test of proportionality — a legitimate aim, a rational connection between the measure and that aim, the absence of a less restrictive alternative, and a proper balance between the right and the public interest pursued.[21]

The DPDP Act is, in large part, the legislative response to Puttaswamy: it is the statutory machinery through which the constitutional right to informational privacy is given operative content in the relationship between individuals and those who process their data. This pedigree matters for how the AI paradox should be resolved. The publicly-available-data exclusion, on its expansive reading, would

[18]Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026 (India), notified February 2026, introducing obligations relating to synthetically generated information and deepfake content.
[19]Ministry of Electronics and Information Technology, India AI Governance Guidelines (Nov. 2025) (non-binding guidance adopting a principles-based, light-touch approach).
[20]Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1 (India).
[21]Id. (articulating the proportionality standard governing intrusions upon informational privacy: a legitimate aim, a rational nexus, the necessity of the measure, and a balancing of the right against the public interest pursued).

permit large-scale processing that the individual never anticipated and cannot control, and would do so without any of the safeguards — notice, purpose limitation, the ability to object — that give the right to informational privacy its practical force. A construction of the statute that produced that result would sit uneasily with the constitutional value the statute exists to vindicate. The interpretive principle that subordinate provisions be read so as to advance, rather than defeat, the object of the parent right therefore counsels against the wide reading of the exclusion, independently of the textual and policy arguments canvassed earlier.

At the same time, Puttaswamy does not support a regime of prohibition. Informational privacy is not an absolute right; it is one to be balanced, through proportionality, against legitimate competing interests, which may include the development of socially valuable technology. A flat ban on training models with any personal data — the result that a literal reading of the consent requirement would produce were the exclusion unavailable — would fail the proportionality test from the opposite direction, restricting a legitimate activity more severely than the protection of privacy requires. The constitutional frame thus pushes against both poles of the paradox: it resists the permissiveness of the wide exclusion and the prohibition of the strict consent requirement, and it points towards a middle path in which model training is permitted on a defined, proportionate basis subject to meaningful safeguards.

There is a further constitutional dimension the present statute under-develops. Puttaswamy's conception of informational privacy is allied to a notion of informational self-determination — the individual's ongoing authority over how data about her is used. The erasure problem identified in Part IV is, in this light, not only a technical embarrassment but a constitutional one: a right of control that the governing technology renders unexercisable is a right hollowed of substance. A constitutionally faithful framework for AI would therefore need to supply alternative mechanisms of control — rights to object to future processing, to suppress particular outputs, and to contest inferences — that can actually be honoured, in place of a formal entitlement that cannot. The constitutional guarantee, properly understood, demands not the rhetoric of control but its reality.

VII. TOWARDS A CALIBRATED FRAMEWORK

If neither prohibition nor permissiveness is acceptable, the task is to design a calibrated middle path. The proposals that follow are offered not as a finished code but as the elements of a framework that would dissolve the paradox while remaining faithful to the statute's purpose and to the constitutional value it serves.

First, the legislature or the rule-making authority should introduce a defined, purpose-bounded lawful basis for the processing of personal data for model training, distinct from both consent and the existing legitimate uses. Unlike the European legitimate-interest basis, it need not be open-ended; it could be confined to training, conditioned on the lawful acquisition of the underlying data, and subject to express safeguards. This would replace the present reliance on the publicly-available-data exclusion with a basis that permits training while keeping the activity within the Act rather than outside it — preserving, crucially, the regulator's jurisdiction over how training is conducted.

Second, the publicly-available-data exclusion should be narrowed and clarified so that it cannot be pressed into service as a general licence for AI training. It should be read, and if necessary amended, to apply only where the data principal herself made the data public and continues to make it available, and it should be made explicit that the subsequent aggregation and model-training use of such data does not inherit the exclusion. This realigns the provision with its evident purpose — sparing the law from policing genuinely and intentionally public information — without allowing it to swallow the statute.

Third, obligations should be graduated by scale. The Act already contemplates the designation of Significant Data Fiduciaries subject to heightened duties; large-scale model developers are natural candidates for such designation.[22] Heightened obligations might include published transparency reports describing the categories and sources of training data, independent audits of data-handling practices, and demonstrable provenance for training corpora. Graduation ensures that the compliance burden falls most heavily where the risk is greatest, and spares genuinely small-scale or research uses from disproportionate cost.

Fourth, the right of erasure should be reconceived for the model context. In place of an unattainable obligation to expunge data from trained parameters, the framework should guarantee operative substitutes: a right to have one's data removed from future training runs, a right to object to the processing of one's data for training at all, and a right to the suppression of model outputs that reproduce or infer one's personal data. These are obligations developers can actually meet, and they deliver the control the constitutional right requires in a form the technology can honour.

Fifth, the research exemption should be preserved and, if anything, clarified, so that genuinely academic and public-interest model development retains a clear and generous pathway distinct from the commercial basis proposed above. The line between research and commercialisation will require

[22]DPDP Act, 2023, § 10 (empowering the Central Government to designate "Significant Data Fiduciaries" subject to enhanced obligations).

careful definition, but the principle — that non-commercial knowledge production deserves lighter-touch treatment — is sound and worth protecting. Taken together, these elements would convert a paradox into a regime: they permit the central activity of the AI economy, but on terms; they keep that activity within the Act's supervisory reach rather than exiling it through an exclusion; and they replace formal rights that cannot be exercised with practical ones that can.

VIII. CONCLUSION

The Digital Personal Data Protection Act, 2023 was built around a single, principled commitment: that individuals should control the use of information about themselves through free, purpose-bound consent. The arrival of large-scale artificial intelligence has tested that commitment against a technology consent cannot easily reach. This paper has argued that the encounter produces a genuine paradox rather than a soluble ambiguity. The consent route is practically closed to the developer of a general-purpose model; the legitimate uses do not fit; and the two exits from the consent requirement fail in opposite directions — the publicly-available-data exclusion too wide to defend, the research exemption too narrow to use. Beneath the question of entry lies the deeper problem of exit: a right of erasure the architecture of trained models cannot honour, exposing the limits of consent as a protective device in this domain.

The resolution lies neither in pretending that the existing exclusion licenses everything nor in reading the consent requirement as a prohibition. It lies in a deliberate, calibrated reform: a purpose-bounded basis for training, a narrowed exclusion, obligations graduated by scale, and rights of control reconceived for what the technology can actually deliver. The constitutional guarantee of informational privacy recognised in Puttaswamy demands as much — neither the hollow permissiveness of unrestricted training nor the futility of an outright ban, but a proportionate framework in which innovation and autonomy are reconciled rather than sacrificed to one another. India has the statute and the constitutional foundation; what remains is the calibration.