DPDP ACT "SPECIFIED PURPOSE" CONSENT VS GENERATIVE AI DATA USE

Inaan Prashere, Punjabi University, Patiala

Emal Biju Pappachan, Punjabi University, Patiala

1. ABSTRACT

This report analyses the tension between India's Digital Personal Data Protection Act, 2023 (DPDP Act) and the data requirements of generative AI systems. While the DPDP Act requires personal data to be processed only with free, informed, specific, and purpose-bound consent, generative AI models rely on large, diverse datasets and continuous training, often making compliance with these requirements difficult. The study identifies key legal and technical conflicts, including the mismatch between AI's broad data use and the Act's purpose limitation, data minimisation obligations, and challenges in deleting or "untraining" data following consent withdrawal.

It also compares the DPDP Act with the EU GDPR and the UK Data Protection Act, highlighting that these frameworks provide broader legal bases for processing than the DPDP Act. The report evaluates potential compliance measures such as the use of genuinely public data, anonymisation, privacy-enhancing technologies, dynamic consent, and stronger data governance, alongside policy reforms including research exemptions and clearer regulatory guidance. It concludes that, without greater flexibility and clearer interpretation of consent requirements, the DPDP Act may hinder AI innovation or result in widespread non-compliance, and recommends practical compliance strategies and policy reforms to better balance privacy protection with AI development.

Keywords: DPDP Act,2023, Generative AI Consent, Specified Purpose, Data Minimization, GDPR, UK Data Protection, AI Compliance, Purpose Limitation, Data Flow, AI Training.

2. INTRODUCTION

India's Digital Personal Data Protection Act, 2023 (DPDP Act)[1] establishes a new, comprehensive privacy law. It focuses primarily on consent as the lawful basis for processing digital personal data, requiring that data be collected and used only for a "specified purpose" communicated to the individual. In parallel, generative AI models (including large language models and image/video generators) have surged, relying on the ingestion and processing of vast, unstructured datasets (often including personal data) to train and improve models.

This raises a pressing question, can AI developers and deployers legally gather and use personal data at scale under India's consent-based regime? The DPDP Act's purpose limitation a core privacy principle seems at odds with the inherently exploratory and iterative nature of AI training. AI systems often learn from large text, images, and user interactions, with purposes that can only be broadly defined (e.g. "train a language model for general knowledge tasks").

Moreover, if consent is required for each specified use of data, continuously updating models with new data may necessitate repeated consent collection. This report conducts a deep dive into this conflict. We analyze the DPDP Act itself and guidance (primary sources), survey academic and industry commentary, and map AI data flows. We also compare how the EU's GDPR, UK law, and US frameworks address similar issues. The goal is to outline the legal constraints, technical challenges, compliance strategies, and enforcement scenarios at the intersection of DPDP's consent doctrine and generative AI data practices.

3. STATEMENT OF PROBLEM

The core issue is that the DPDP Act mandates personal data processing only with the principal's consent for a specifically notified purpose[2] (section 6(1)) Generative AI development, by contrast, typically involves collecting and processing vast amounts of data for purposes that are not narrowly defined and may evolve over time. Key problems include:

- **Purpose Specificity vs. AI Flexibility:** How can an AI developer obtain "specified purpose" consent for broad or unpredictable uses of data (e.g. training on text to power various applications)?

[1] The Digital Personal Data Protection Act, No. 22 of 2023, India Code (2023).
[2] DPDP Act, § 6(1).

- **Data Minimization vs. Data Scale:** DPDP requires collecting only necessary data for the purpose, but AI models often benefit from using any available data to improve accuracy (data minimization is technologically challenging).

- **Withdrawal of Consent:** If a user withdraws consent, models may have already "absorbed" their data, raising questions about deletion/unlearning. DPDP requires ceasing processing on withdrawal[3].

- **Continuous Learning:** Many AI systems continuously ingest new data. Repeated consent collection for ongoing training seems impractical.

- **Scope of Legitimate Exceptions:** DPDP allows some "legitimate uses" without consent (e.g. government functions, emergencies[4]). Do these cover AI training? If not, must all training rely on consent?

- **Publicly Available Data:** The Act exempts data made publicly available by the data principal (e.g. their public social media). Does this permit scraping web data for AI? The definition is narrow, applying only if the individual or law made it public, unlike "public domain" broadly. Addressing these tensions is crucial for both protecting data principals and enabling AI innovation. This research aims to systematically evaluate these issues.

4. RESEARCH QUESTIONS

1. What does the DPDP Act require for valid consent and "specified purpose" under its consent doctrine?

2. How do generative AI development and deployment processes (training, fine-tuning, inference, continuous learning) handle personal data, and where do they clash with DPDP consent rules?

3. What lawful bases or compliance mechanisms could allow generative AI models to operate within India's DPDP framework (e.g. consent, anonymization, research exceptions, contractual

[3] Id. § 6(4).
[4] Id. § 7.

arrangements)?

4. How do other jurisdictions (EU GDPR, UK Data Protection Act, US laws/policies) address similar issues, and what lessons or analogies apply?

5. What enforcement risks and litigation scenarios arise if AI developers use personal data without meeting DPDP requirements?

5. OBJECTIVES

- **Analyze DPDP Consent Doctrine:** Examine the text of the Act and official rules to clarify how consent, withdrawal, purpose limitation and exemptions work.
- **Map AI Data Flows:** Describe generative AI's data lifecycle (data collection, training, model use, updates) and identify points of friction with DPDP.
- **Identify Legal Tensions:** Pinpoint conflicts between DPDP's purpose limitation and AI's data needs, including issues around continuous data ingestion and model reuse.
- **Evaluate Compliance Strategies:** Explore possible lawful bases (e.g. contractual consent, legitimate uses, anonymization) and technical solutions (differential privacy, synthetic data, DPIAs) for AI to comply.
- **Comparative Analysis:** Illustrate how GDPR/ UK/US handle purpose and AI data, drawing parallels and contrasts.
- **Policy Recommendations:** Propose how Indian regulators and industry can reconcile DPDP with AI, and outline potential enforcement and remedies scenarios.

6. RESEARCH METHODOLOGY

This research uses doctrinal legal analysis of primary sources (DPDP Act text, official rules, and guidelines). We review technical literature on generative AI architectures to understand data requirements. A comparative legal approach examines GDPR, UK Data Protection Act, and US privacy laws/policies to draw instructive analogies. Relevant academic and industry commentary (law review articles, think-tank reports, firm whitepapers) is surveyed to capture existing analyses.

We synthesize these strands into structured sections, supported by citations to authoritative sources (legislation, regulator guidance, and peer commentary). Tables and diagrams are used to clarify data flows, regulatory milestones, and comparative provisions. Throughout, emphasis is on reconciling technical AI needs with legal constraints, noting gaps or ambiguities that may require policy action.

7. LITERATURE REVIEW

The interplay between AI and data protection has attracted recent analysis. Several authors note that India's DPDP Act is consent-centric, lacking broad lawful bases like GDPR's legitimate interest. Mohamed (FPF, 2024) observes that DPDP's reliance on consent (with only narrow "legitimate uses" exceptions) makes traditional AI training (often using scraped data) challenging[5]. Summers & Thompson (ReedSmith, 2023) similarly explain that DPDP processing is only allowed by specific consent or narrow statutory [6]uses.

They stress that notices must itemize data types and specific purposes, reinforcing the purpose limitation. More technically oriented works highlight AI data issues. Jha et al. (2026) explicitly warn that DPDP's consent model is "functionally unworkable" for billion-token AI training datasets, given the weak ties between individuals and such aggregated [7]data. Singh & Thakur (2025) map AI data lifecycle under DPDP, noting extraterritorial scope and the carve-out for self-posted public data[8]. They emphasize that consent must be specific and easily withdrawable, influencing how feature stores and training pipelines are designed.

Other experts (e.g. Kemp 2020; Latham & Watkins 2023) have compared DPDP to GDPR, noting DPDP's stricter data minimization and purpose rules[9]. At the regulatory level, the UK ICO has published guidance on purpose limitation in AI, stressing separate documentation of training vs application purposes[10]. In sum, scholarship consistently identifies a tension:

[5] B. Mohamed, Five Ways in Which the DPDPA Could Shape the Development of AI in India, Future of Privacy Forum (2024).
[6] M.S. Summers & T.J. Thompson, India in Focus: Data Protection and AI in India, Reed Smith Blog (2023).
[7] A.K. Jha, M. Jain & T. Bhawsar, Intellectual Property and Personal Data in AI Datasets under India's DPDP Act, 4 Trends IP Res. 45, 48–50 (2026).
[8] A. Singh & S. Thakur, Data Protection and AI under the DPDP Act, 2023: An Indian Perspective, 13 Int'l J. Creative Res. Thoughts 105, 108–12 (2025).
[9] See R. Kemp, Algo IP: Intellectual Property in AI Datasets, Kemp IT Law (2020); Latham & Watkins, India's DPDP Act vs the GDPR: A Comparison (Whitepaper, 2023).
[10] Info. Comm'r's Off. (UK), Generative AI: Purpose Limitation Consultation (2023); see also Info. Comm'r's Off. (UK), Guidance on the Data Protection Principles (2026).

DPDP's formal consent rules struggle to accommodate AI's broad, iterative data use model.

8. LEGAL FRAMEWORK UNDER DPDP ACT

8.1 Scope and Definitions

The DPDP Act (2023) applies to "digital personal data" of Indian individuals (Data Principals) and to any processing (even abroad) in connection with offering goods or services to Indians. Key definitions include Data Fiduciary (entity determining processing purpose) and Data Principal (individual whose data is processed). Notably, the Act does not create special categories of sensitive data, all personal data is treated similarly.

The Act excludes certain data from its ambit, personal data processed for personal/domestic use, or data made publicly available by the individual or by law[11]. For example, if a person publicly posts personal information on social media, that data falls outside DPDP regulation. This carve-out is narrower than a general "public domain" exemption and does not cover personal data made public by others (e.g. news articles about someone).

8.2 Consent and Notice (Sections 5–6)

Consent is central. Section 5(1) requires that any request for consent be accompanied by a notice containing, at minimum: (i) the personal data types and the purpose of processing; (ii) the data principal's withdrawal and rights procedures; and (iii) complaint channels[12]. The notice must be clear, standalone, and in plain language. Section 6(1) then mandates that consent be "free, specific, informed, unconditional and unambiguous", given via a clear affirmative action[13]. Crucially, consent must be tied to a specified purpose and limited to data necessary for that purpose.

For example, a consent for telemedicine can only cover health-related data, not unrelated contact lists. Consent can be withdrawn anytime (Section 6(4)), and on withdrawal the fiduciary must stop processing "as soon as reasonably practicable" unless another lawful basis exists[14]. By design, this regime is consent-centric. There is no provision for processing on "legitimate interests" (as under EU law) or contract necessity. All consent terms must stand

[11] DPDP Act, § 3(c).
[12] Id. § 5(1).
[13] Id. § 6(1).
[14] Id. § 6(4).

independently (waivers of rights are void) and be granular. The Act even contemplates "Consent Managers" to help data principals give and manage their consents [15](Section 6(7)-(9)), underscoring the emphasis on individual control. In short, processing is only lawful if it aligns strictly with the consented purpose and declared data scope.

8.3 Legitimate Uses (Section 7)

The Act provides a closed list of "certain legitimate uses" (Section 7) where consent is not required[16]. These include:

- **(a)** Uses for the very purpose for which the DP voluntarily provided the data and did not object to its use (e.g. a customer gives address to a seller, the seller may use it to send the purchase receipt).

- **(b)** Uses by the State or its agencies to deliver specified public services (e.g. welfare benefits), where the DP either consented to that use or the data is already in a notified government database.

- **(c–h)** Statutory obligations, legal compliance, emergencies (medical or public health threats), disaster relief, employment-related security, etc.

These exceptions are quite limited. Significantly, none explicitly cover commercial AI training or corporate analytics (outside emergency or statutory compliance scenarios). Thus, outside these narrow cases, even sophisticated machine learning may only proceed if consent is validly obtained.

8.4 Data Principal Rights

Data principals have rights including information, correction/updating, withdrawal of consent, and grievance redressal. Withdrawal of consent must be "as easy as giving it". Upon withdrawal, fiducialies must erase or anonymize the data unless retention is required by law[17] (Section 8(7)). The Act envisages a Data Protection Board of India to adjudicate complaints and impose penalties.

[15] Id. § 6(7)–(9).
[16] Id. § 7.
[17] Id. § 8(7).

8.5 Enforcement and Penalties

Chapter VIII empowers the Data Protection Board to investigate breaches (upon complaint or suo moto), order cessation of processing, and levy monetary penalties. Breach of the law can attract very high fines (the schedule authorizes up to Rs. 50 crore or even 2500 crore depending on the violation). Repeated offences can double penalties. The Board can also require mandatory audits and impact assessments for "significant data fiduciaries" (Section 10). Notably[18], unlike some laws, DPDP does not explicitly provide a private right of compensation to individuals (no statutory cause of action), but individuals can complain to the Board for enforcement relief. The emphasis is on regulatory enforcement, with heavy deterrent fines for non-compliance.

9. "Specified Purpose" Consent under DPDP

The term "specified purpose" is defined as the purpose "mentioned in the notice" given to the principal. In practice, this means the fiduciary must describe concretely what it intends to do with the data (e.g. "to provide online medical consultations"). The consent then authorizes processing only for that purpose. Any attempt to use the data for another purpose (even if related) would lack valid consent. The rules further clarify that the notice should include an itemized description of the goods/services or uses enabled by the processing. In effect, DPDP demands a one-to-one link between consent and purpose.

Key interpretative points: The Act does not specify whether multiple purposes can be bundled in one consent; the language ("the purpose" and "specified purpose") suggests a primary purpose is required. In practice, if data will be used for multiple related objectives, a compliant approach would be to enumerate all those purposes clearly in the notice. Consent must cover exactly that list, any extraneous use (not explicitly stated) would be unauthorized.

For example, a fintech app asking consent to "process financial data for loan eligibility" could not later use that data for unrelated marketing unless it had separately obtained consent.

This narrow interpretation is underscored by the data minimization requirement, data collected must be "necessary for such specified purpose". Unlike some laws that allow secondary uses compatible with the original purpose, DPDP's regime implies that any new use likely requires

[18] Id. § 10.

fresh consent or fall under an exception (e.g. emergency). The UK ICO's guidance similarly emphasizes that each distinct processing activity in the AI lifecycle (training vs inference) should have its own clear purpose[19]. Finally, the DPDP rules (draft 2025) reinforce transparency the notice must stand alone, be understandable without other documents, and provide withdrawal/rights mechanisms. This shows regulators expect very granular purpose descriptions, not vague or bundled statements. In sum, "specified purpose" under DPDP is intended to be strict and narrow, aligning data use strictly with what users consented to.

10.GENERATIVE AI DATA NEEDS AND FLOWS

Generative AI models (e.g. GPT-type language models, image generators) are trained on large unstructured datasets (text, image collections, audio, etc.) to learn patterns for creating new content. The typical data lifecycle involves:

- **Data Collection:** Aggregating massive amounts of raw data. This may include web crawling (news articles, blogs, forums, social media), user-provided content, proprietary datasets, sensor streams, etc. The data can be heterogeneous (text, images, audio, code). Not all data is personal, but often personal data is incidentally included.

- **Data Preprocessing:** Cleaning and formatting data (tokenization for text, labeling for supervised tasks). This stage often filters out obvious PII, but large models may still encode identifiable text snippets.

- **Model Training:** Using the data to train a core (foundation) model. This typically requires an enormous dataset (billions of tokens) and considerable compute. The purpose of initial training is to learn general language or visual patterns, not tied to a specific end-use application.

- **Fine-Tuning:** Adapting the core model to a narrower application by training on a smaller, domain-specific dataset (e.g. medical texts for a health assistant). This is a separate processing purpose.

- **Inference/Deployment:** Using the trained model to generate outputs (e.g. answering questions). The model may continue to learn from user feedback (sometimes called

[19] Info. Comm'r's Off. (UK), Generative AI: Purpose Limitation Consultation (2023).

"online learning" or reinforcement learning from user interaction).

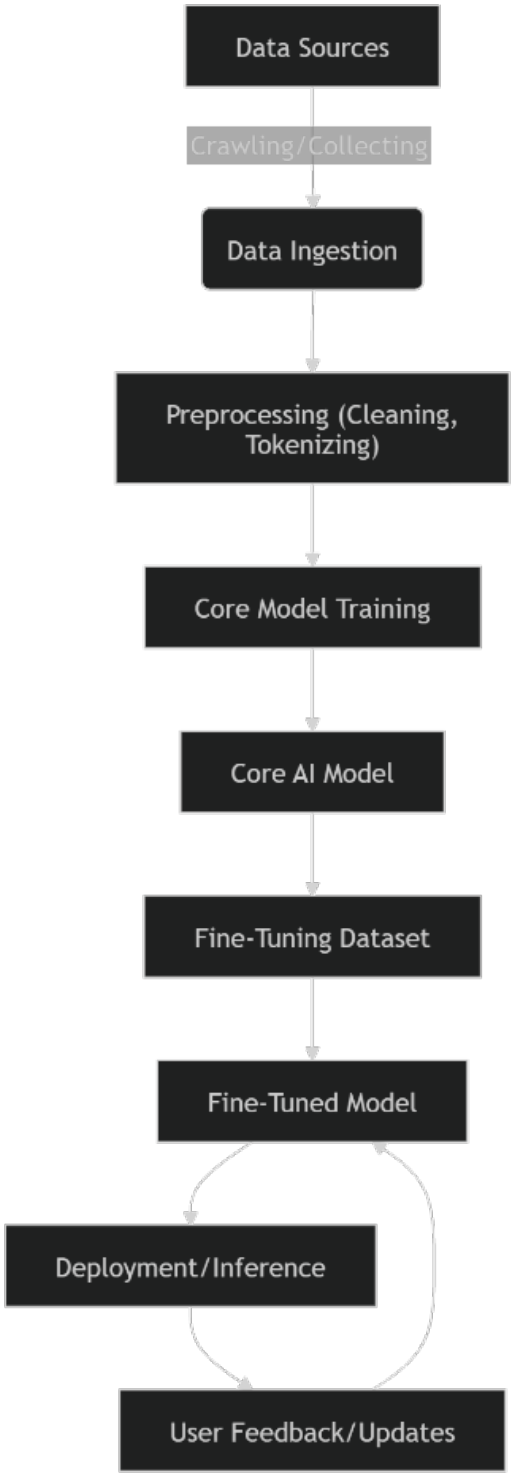

Figure 1: Simplified data flow in generative AI model development (source: ICO analysis and model engineering knowledge). Data from varied sources is ingested, preprocessed, and used to train a core model, which may then be fine-tuned and deployed. User feedback can loop back into model updates.

Each stage can involve personal data. The ICO notes that AI has separate purposes at each stage (core model training vs application use). For instance, collecting general web text to train a model is a different purpose than later using that model to power a chatbot. Data minimization is challenging here as model performance often improves with more diverse data [20](Kaplan & McCandlish 2020). Erasing individual data points from a model is technically difficult (the "unlearning" problem). Moreover, generative models may inadvertently memorize and reproduce sensitive data (as seen in some LLMs). In summary, generative AI needs large-scale, iterative data ingestion with often diffuse or evolving purposes. These characteristics directly clash with DPDP's strict purpose, scope, and minimization norms.

11. POINTS OF CONFLICT

Several tensions emerge when overlaying the DPDP Act onto this AI data flow:

- **Broad vs Specific Purpose:** AI training often has a general purpose (e.g. "train a language model") that can serve many downstream uses. DPDP requires specific notice of purpose, making it hard to anticipate all future applications. If an AI developer specifies "training a sentiment analysis model" as purpose, using the same data to also train a translation model would lack consent. Essentially, once a model is trained, a wide range of uses may arise, potentially conflicting with the original consent scope. The ICO warns that one core model giving rise to "many different applications" requires careful delineation of purposes[21].

- **Data Minimization vs. Large Data Needs:** AI performance generally improves with more data. However, DPDP section 6(1) mandates collecting only data "necessary for the specified purpose[22]". It is debatable whether any individual piece of data is strictly necessary for a broad training task. If a Data Principal gave consent for "training a model," does that automatically allow all data from them? Likely only if explicitly covered by the purpose. This contradicts the usual AI approach of collecting all available relevant data. Models often use information far beyond what was foreseeable at consent time.

[20] See J. Kaplan et al., Scaling Laws for Neural Language Models, arXiv preprint arXiv:2001.08361 (2020).
[21] Info. Comm'r's Off. (UK), Generative AI: Purpose Limitation Consultation (2023).
[22] DPDP Act, § 6(1).

- **Continuous vs One-Time Consent:** Many AI systems continuously learn from new data (e.g. user interactions, new web content). DPDP consent is ordinarily one-off, with withdrawal possible later. Should new data ingestion require fresh consent each time? If consent was initially for a fixed dataset, adding new data would arguably need new notice/consent. Continuous re-consenting at web scale is impractical, suggesting structural conflict.

- **Withdrawal of Consent:** If a data principal later withdraws consent, DPDP requires the fiduciary to cease processing and erase data as soon as possible[23]. In AI terms, this would mean removing the individual's data from the model and training dataset. Current AI technology does not easily permit "forgetting" data once integrated into model parameters. This creates a potential compliance gap: how to render a model free of withdrawn data? At minimum, it necessitates designing pipelines to track and purge data, which is non-trivial. Arshpreet and Thakur (2025) note that withdrawal "influences feature stores, cached embeddings, and backup pipelines" in AI systems[24].

- **Data Principal Relationship:** DPDP is predicated on a direct (or at least identifiable) relationship between fiduciary and principal. In generative AI, training data often comes from sources (forums, books, etc.) with no direct relationship to each individual DP. Jha et al. observe that in massive corpora, individual data-principal links are "attenuated," making it unclear how meaningful consent can even be obtained or managed at scale[25]. If the DP has never interacted with the fiduciary, obtaining valid "free, informed" consent is very difficult.

- **Publicly Available Data Use:** A bright spot is the Act's narrow exemption for public data posted by the DP[26]. In principle, an AI developer could rely on scraping a person's own public blog or social media posts without consent. However, the exemption is limited, data must be "publicly made available by the individual" (or by law). Content about the person on news sites or others' blogs would not qualify. This narrowness may cause confusion, many AI datasets include scraped content not explicitly posted by those individuals. Unlike GDPR (which treats all publicly available personal data as

[23] Id. § 6(4), § 8(7).
[24] Singh & Thakur, supra note 6, at 110.
[25] Jha et al., supra note 5, at 49.
[26] DPDP Act, § 3(c)(b).

still regulated), DPDP effectively de-regulates personally posted public data. For example, a user's Twitter posts (if the user is considered as posting their personal data publicly) would fall outside DPDP, potentially allowing indexing and use. But scraping profiles or contact info added by others (even if on a public forum) remains regulated. This could encourage AI developers to rely heavily on user-generated public content, but with uncertainty at the margins.

- **No Legitimate Interest Basis:** Unlike GDPR, DPDP has no catch-all lawful basis like "legitimate interests". For GDPR, AI firms often argue that model training is within legitimate interests (after a balancing test) or for scientific research (with safeguards). Under DPDP, absent a consent or a narrow Section 7 exception, there is no parallel fallback. This restricts options for cases where obtaining consent is impractical. The only potential leanings are "certain legitimate uses" (e.g. if an AI project qualifies as a government health study, possibly exempted) or future "research" rules (see below). Without legislative change, generative AI firms in India will struggle to find lawful bases beyond explicit consent.

- **Children's Data:** Section 9 adds extra hurdles for children's data (requiring parental consent, banning targeted ads, etc[27].). Generative AI scrapers often ingest content from and about minors inadvertently. Use of any child data (as defined) would require stringent protections. The Act even permits a government to lift some child-data rules for vetted fiduciaries. Nonetheless, any training on data of minors would face heavy compliance requirements.

These conflicts are not merely theoretical. As one analyst notes, DPDP's architecture is "functionally unworkable" for the scale of AI corpora in practice. Without mitigation, a large language model used in India today (scraping billions of web documents) would be technically in violation of DPDP unless explicit consent was obtained from every Indian individual whose data appears. This motivates exploring potential workarounds and international comparisons.

12. COMPLIANCE STRATEGIES AND OPTIONS

Given these challenges, we consider ways AI developers might seek compliance or mitigation

[27] Id. § 9.

under DPDP:

- **Explicit Consent Processes:** The most straightforward (if onerous) approach is obtaining valid consent from data principals for AI use. For example, a platform could ask users to agree that their data may be used to improve AI models, specifying purpose and data categories. Consent should ideally be fine-grained (e.g. separate consents for different data types or uses). Use of Consent Managers (per Section 6) or privacy dashboards might help manage these consents. However, for public web scraping (no direct user interface), this is impractical, it might only work for data collected in-app or via APIs from consenting users.

- **Legitimate Uses Claimed Narrowly:** Section 7's exceptions could be invoked in limited scenarios. For example, a government AI research initiative might qualify under (b)/(c) (public interest or government function). An AI tool for emergency response could cite the medical or disaster exceptions (f–h). But these are fact-specific and likely inapplicable to commercial AI developers. The workplace exception (i) might cover AI used internally for employee safety or trade secret protection. In practice, most generative AI tasks (e.g. chatbots, content generation) do not neatly fall under any Section 7 category, so legitimate uses provide little flexibility for non-public actors.

- **Anonymization and Pseudonymization:** If data can be effectively anonymized, it falls outside the Act altogether. AI firms could anonymize personal data before use (e.g. remove names, identify fields). Fully anonymized datasets (with no re-identification risk) would not be "personal" data. However, truly robust anonymization for unstructured data (like free text) is very hard, models can often re-generate personal snippets if trained. Pseudonymization (masking direct identifiers) still counts as personal data under DPDP. Still, strong de-identification and privacy-enhancing technologies (PETs) like differential privacy could mitigate risk. Reporting compliance with DPDP might involve demonstrating such PET usage. But regulators may still question if the data is really non-personal.

- **Research Exemption:** DPDP (draft Rule 17(2) (b)) contemplates exemptions for research/ archiving/statistical use if certain standards are met. This resembles GDPR's research exemption but is currently undeveloped (no notified standards yet). If eventually implemented, it could allow AI researchers to use personal data without

consent, subject to safeguards (e.g. IRB oversight, data security). Firms might anticipate such rules to facilitate AI development in "research mode." For now, this remains speculative.

- **Dynamic Consent and Ongoing Notices:** Some have proposed dynamic consent models (via apps or platforms) where users are continuously informed and can opt in/ out over time. For example, a social media company could allow users to opt into having their posts used to train new AI features. Each new feature rollout could trigger a consent update. This aligns with DPDP's emphasis on ongoing control. However, it requires robust systems to track consents, and only covers data collected with direct user relationships, not open web data.

- **Contractual Arrangements:** With no contractual necessity ground in DPDP, one could consider embedding consent-like clauses into user contracts (e.g. terms of service). But Section 6 expressly forbids consent inferred from silence or broad agreements, active affirmative consent is required. Thus "forced consent" via TOS might not satisfy the "clear action" standard. Nonetheless, companies might combine privacy notices with TOS mechanisms (as long as they don't waive rights).

- **Combined Approaches:** Some hybrid models are emerging. For example, LinkedIn's AI research director has called for licensing models of data use (like collectively licensing web data) or requiring AI firms to pay for data access under various IP regimes (copyright, trade secrets). While not a DPDP solution per se, these approaches could change how personal content is used (e.g. platforms might only allow scraping via paid APIs, with their own terms). Similarly, the Indian DPIIT's AI working paper suggests voluntary data standards and aggregate anonymized data pools for AI. Regulators could promote industry frameworks (data trusts, model registries) that incorporate DPDP principles (logging consents, bias audits, etc.).

Strategy	**Advantages**	**Drawbacks**
Obtaining Explicit Consent	Aligns with the law, strongest legal basis. Allows clear tracking of purposes.	Impractical at AI scale. high user friction as obtaining consent for all data is impractical.
Narrow Legitimate Use	No consent needed if the	Extremely limited scope, not

	statutory criteria are met (e.g., government use).	suitable for most private AI projects.
Anonymization/Privacy Enhancing Technologies (PETs)	Personal data is no longer regulated if it is truly anonymous.	Difficult to fully anonymize large datasets, risk of re-identification as the DPDP Act still requires reasonable security measures.
Research Exemption	Could enable broader use of data for research and development (R&D) with appropriate safeguards.	Provisions remain unspecified and are pending implementation; applicability to corporate AI is uncertain.
Dynamic Consent	Keeps users informed and provides flexibility in managing consent.	Complex to implement as it requires an active user base, difficult to retrofit for web-scraped data.
Licensing / Data Pool	Centralises data use and can enforce usage rules (e.g., through commercial APIs).	Depends on third-party cooperation and may still require individual consent.
Differential Privacy	Reduces privacy risks and is viewed favourably by regulators.	Involves a privacy–utility trade-off and may reduce AI model performance.

Table: Compliance options vs pros/cons for generative AI under DPDP. (Sources: DPDP Act; Mohamed 2024; Singh & Thakur 2025)

13. COMPARATIVE PERSPECTIVES (EU/UK/ US)

Looking beyond India, data protection laws grapple with AI's data needs differently:

- **EU GDPR:** Like DPDP, GDPR enshrines purpose limitation [28](Art. 5(1)(b) requires specified, explicit, legitimate purposes) and data minimization [29](Art.5(1)(c): "adequate, relevant and limited to what is necessary"). However, GDPR offers multiple legal bases (Art.6): consent, contractual necessity, legal obligation, vital interests, public interest, and notably, legitimate interests[30]. AI training often relies on legitimate

[28] 28. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation), art. 5(1)(b), 2016 O.J. (L 119) 1 [hereinafter GDPR].
[29] Id. art. 5(1)(c).
[30] Id. art. 6(1)(f).

interest, following a balancing test (Recital 47). GDPR also has explicit research exemptions, further processing for scientific research is often permitted under certain safeguards [31](Art.5).

- Importantly, GDPR does not exempt publicly available data from regulation, any personal data even if public, is still subject to GDPR if used (subject to the same principles). Thus, scraped web data is generally regulated. Some EU authorities have warned that large-scale text mining may violate GDPR unless consent is obtained or research grounds apply (e.g. EDPS guidance on generative AI).

- **UK Data Protection Act (UK GDPR):** The UK regime is similar to EU law[32]. The UK's ICO has specifically issued draft guidance on purpose limitation in AI, emphasizing that training data reuse requires compatibility checks or new legal basis[33]. The ICO also treats legitimate interests as a possible lawful basis for training, subject to justification. The US-UK Data Bridge (from DPDP Board minutes) suggests the UK will take a more risk-based approach to AI, but no change to consent doctrine is planned. The UK also has no "public data" exemption akin to India's; what matters is whether individuals are identifiable.

- **United States:** The U.S. lacks a comprehensive federal privacy law. Instead Sectoral laws for sensitive data like health (HIPAA) or children's (COPPA) have special consent rules [34](COPPA requires verifiable parental consent for minors under 13). Some states (e.g. California's CCPA/CPRA) have broad scope and require notice of purpose and opt-out for sale of personal data[35]. California law requires that data be "collected, processed, retained, etc. only for specified, explicit and legitimate purposes", echoing GDPR's language, and limits further use beyond consumer expectations. However, the CCPA does not require consent for general data collection (it's mostly about notice and opt-out of sales) and has no system of lawful bases aside from consent for use of sensitive data. CPRA (2023) adds some right to limit use of sensitive personal info. In general, U.S. companies building AI often rely on terms of service and claims of fair use (in copyright) or rely on publicly available data under the First Amendment (for

[31] Id. art. 5(1)(b), art. 89.
[32] Data Protection Act 2018, c. 12 (UK).
[33] Info. Comm'r's Off. (UK), Generative AI: Purpose Limitation Consultation (2023).
[34] Children's Online Privacy Protection Act of 1998, 15 U.S.C. §§ 6501–6506.
[35] Cal. Civ. Code §§ 1798.100–1798.199.100 (West 2023).

speech) but a clear U.S. regulatory approach to AI data is still evolving (see recent White House AI Executive Order).

Various guidelines (NIST AI Risk Management Framework, OECD Principles) emphasize transparency and privacy, but are non-binding. The proposed EU AI Act (regulating AI risk levels) and the U.K.'s AI Safety Institute don't directly amend data laws, but a strict DPDP may effectively act as a de facto data-use regulation for AI in India. In summary, EU/UK laws share DPDP's purpose limitation ethos but provide more flexibility via other lawful bases or research exemptions, potentially making AI compliance somewhat easier. U.S. law's piecemeal approach means companies may face different constraints (e.g. California opt-outs) but no uniform consent requirement like DPDP. Indian firms and regulators may look to GDPR/ICO for guidance on applying principles like data minimization and compatibility assessments to AI (as the ICO consultation does).

14. ENFORCEMENT RISKS AND REMEDIES

Under DPDP, non-compliance by AI developers could trigger enforcement actions. Affected data principals can complain to the Data Protection Board, which has investigatory and punitive [36]powers. Potential violations include processing without valid consent or for an undisclosed purpose. Penalties are severe (tens to thousands of crores of rupees) for "significant" breaches[37]. For example, if an AI company scrapes personal data without notice or consent, the Board could impose fines and mandate deletion of data. Repeated or egregious breaches could double fines. Besides fines, the Board may order corrective measures: cease processing, implement DPIAs, or require proof of anonymization practices. It can also block access to data for certain entities (similar to GDPR's Article 58 powers). However, unlike GDPR, the Act does not explicitly provide for individual damages claims, so victims would rely on regulatory relief or file in civil court under other laws (e.g. a breach of confidentiality tort). The Board could also invalidate dubious consents [38](Section 26(2) suggests invalidating any contract term that waives rights).

In practice, enforcement of AI-related privacy breaches could mirror ongoing AI litigation trends. For instance, globally there are lawsuits against AI firms for unlicensed data use (e.g.

[36] DPDP Act, §§ 27–28.
[37] Id. § 33, sched. 1.
[38] Id. § 26(2).

authors alleging copyright infringement). In India, one could envision writ petitions or class actions arguing that "specified purpose" was violated. A data principal could allege that using their data to train a public AI chatbot (not consented to) violates Sections 5 and 6. The regulator might investigate whether sufficient notice was given and whether withdrawal rights were upheld. Risk is higher for "significant data fiduciaries" (those dealing with large volumes of sensitive data), these entities face mandatory DPIAs (Section 13) and are under closer scrutiny. AI companies in India should thus treat DPDP compliance as a key governance issue. Non-compliance not only risks fines, but could undermine public trust and lead to bans (e.g. the Board can recommend blocking of non-compliant services). Overall, the enforcement environment is stringent, reflecting the Act's strong protectionist stance.

15. FINDINGS

- **DPDP Act mandates narrow, purpose-tied consent:** The Act's provisions clearly require consent to be specific to a declared purpose and limit data to what is necessary. Notice must describe data and purpose in detail. This formalistic regime contrasts with AI's broad, evolving use of data.

- **Generative AI pipelines generally conflict with DPDP norms:** The nature of AI training large corpus ingestion for general learning does not fit neatly into a single specified purpose. AI workflows involving continuous updating and re-use of data run counter to DPDP's purpose limitation and data minimization requirements. Notably, scholars conclude DPDP's consent model is "unworkable" at AI scale.

- **Limited lawful bases under DPDP:** Without legitimate interest or broad research exemptions, AI developers have few statutory leeways besides consent. The "legitimate uses" exceptions are unlikely to cover most AI activities (except certain government/ public health projects). The carve-out for self-posted public data helps only partly (it excludes third-party data).

- **Comparative regimes offer more flexibility:** GDPR/UK permit legitimate interest and research bases, easing AI training compliance (though still challenging). US approaches vary by sector/state but generally rely on notice/opt-out rather than strict purpose fidelity. Indian firms may not find domestic precedents for large-scale AI data use under DPDP.

- **Compliance requires complex solutions:** Mere legal compliance (seeking consent for every use) is operationally hard. Alternative strategies (anonymization, dynamic consents, data licenses) are being explored but have tradeoffs. For now, many AI deployments in India may face a compliance gap unless either the law or practices adapt. Regulatory uncertainty remains as the DPDP Rules (draft) and future guidelines will influence outcomes (e.g. how "certain legitimate uses" are interpreted, or if research exceptions are fleshed out). In the absence of clear regulatory guidance specifically for AI, companies must be on the side of stringent consent and data hygiene.

16. RECOMMENDATIONS

- **Regulatory Guidance:** The Data Protection Board (or MeitY) should issue guidance on AI. Clarify how "specified purpose" may be defined for AI (e.g. allow layered consents), and how model reuse aligns with purpose. Publish standards for research exemptions promptly to facilitate legitimate AI R&D.

- **Consent Architecture:** Firms should build granular consent systems. Use "layered notices" and clear dashboards (the rules encourage clear, standalone notices). For AI features, seek fresh consents tied to each model version. Consider dedicated Consent Managers to track user permissions over time.

- **Data Governance Measures:** Adopt Privacy-Enhancing Technologies (differential privacy, federated learning) to reduce personal data exposure. Maintain audit trails of data provenance and consent status. Perform DPDP-compliant DPIAs for any high-risk AI (likely mandatory for large models).

- **Use Exempt Data Strategically:** Maximize use of public-domain or company-collected data (with consent). For web data, prioritize content that clearly falls under the "publicly made available by DP" exemption (e.g. voluntarily posted content). License data when possible (negotiated data sets) to avoid reliance on scraped personal info.

- **Advocacy and Policy Engagement:** Industry bodies should engage with regulators on AI-data policy. Propose balanced frameworks: for example, an AI-specific data-sharing

license scheme or codes of practice that respect DPDP aims. Monitor international approaches (EU AI Act, US AI guidelines) for evolving norms.

- **Training and Awareness:** Educate data scientists about DPDP requirements. Encourage "privacy by design" in AI projects, i.e. build models knowing that some data might need deletion or never be collected.

- **Risk Assessment:** Regularly review AI use cases for privacy risk. Prioritize obtaining consent where needed or altering the model to avoid personal data (e.g. filter out names from training). Be prepared to demonstrate measures to the Data Protection Board.

17. CONCLUSION

India's DPDP Act introduces robust privacy protections with a strong consent and purpose-limitation framework. While this upholds individual autonomy, it poses significant compliance challenges for generative AI, which thrives on broad, continuous use of diverse data. Our analysis shows that under the Act's current form, many typical AI data practices would run afoul of the law unless carefully managed. Bridging this gap requires both legal and technical innovation like evolving consent mechanisms, rigorous anonymization, and perhaps most importantly, policy evolution. The DPDP Act's strict approach signals India's commitment to privacy, but achieving a workable balance with AI will likely need clarifications (e.g. how multiple purposes are consented) and possibly legislative updates (e.g. expanded research uses or clarified public data rules).

Meanwhile, AI developers targeting the Indian market must proactively align with DPDP by obtaining explicit consent, minimizing personal data in models, and engaging with regulators. In comparative perspective, India's regimen is more rigid than the GDPR/UK systems (which offer some flexibility) and significantly more structured than the patchwork of US privacy laws. As both AI innovation and data protection evolve globally, India's model may serve as a test case of strict consent-based governance. Ultimately, fostering responsible AI will depend on collaborative approaches, industry self-regulation, clear guidance from regulators, and possibly new statutory exceptions tailored to AI's realities.

BIBLIOGRAPHY

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

2. Info. Comm'r's Off. (UK), Guidance on the Data Protection Principles (2026).

3. R. Kemp*, Algo IP: Intellectual Property in AI Datasets*, Kemp IT Law (2020).

4. Info. Comm'r's Off. (UK), Generative AI: Purpose Limitation Consultation (2023).

5 .Latham & Watkins, India's DPDP Act vs the GDPR: A Comparison (Whitepaper, 2023).

6. M. S. Summers & T. J. Thompson, India in Focus: Data Protection and AI in India, Reed Smith Blog (2023).

7. B. Mohamed, *Five Ways in Which the DPDPA Could Shape the Development of AI in India*, Future of Privacy Forum (2024).

8. A. Singh & S. Thakur*, Data Protection and AI under the DPDP Act, 2023: An Indian Perspective*, 13 Int'l J. Creative Res. Thoughts 105 (2025).

9. A. K. Jha, M. Jain & T. Bhawsar, *Intellectual Property and Personal Data in AI Datasets under India's DPDP Act*, 4 Trends IP Res. 45 (2026).