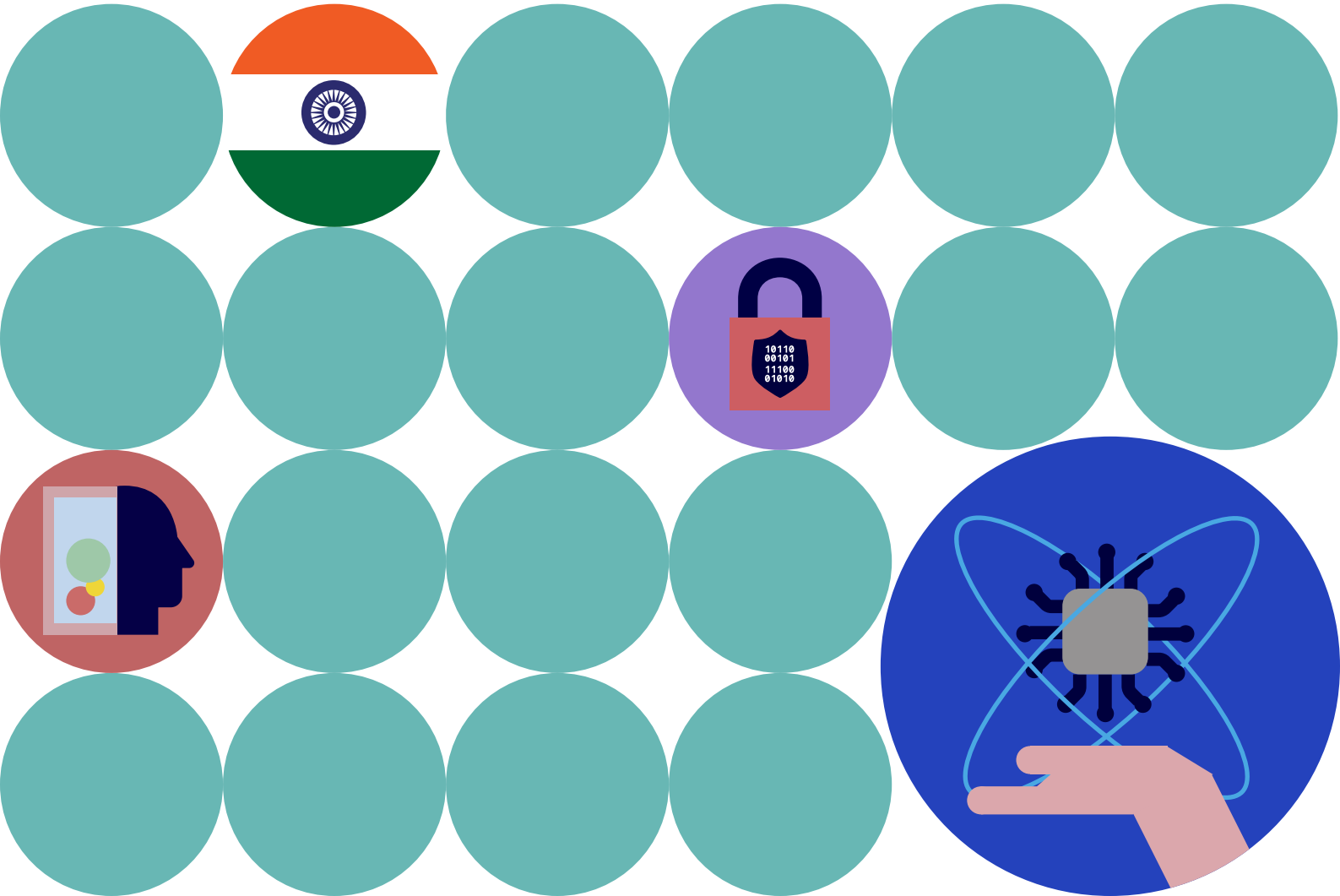
Open Loop | Meta

AI Innovation, Effective Anonymization & the DPDP Act



Contents

Foreword 3

→ **Executive Summary** 7

1 **Introduction** 10

2 **Findings and observations** 15

3 **Recommendations** 28

Recommendation A:
Issue an exemption for AI model training under Section 17(5) 29

Recommendation B:
Issue clear guidance on anonymization for AI development 32

Recommendation C:
Issue guidance on privacy-enhancing technologies (PETs) where appropriate 34

4 **Conclusion** 36

References 38

Annex 1: Glossary (Key Concepts) 40

Annex 2: Research Methodology 42

Annex 3: Policy Background 44

(1) Notable Provisions of DPDP Act and Rules Framework 44

(2) India's AI Policy Landscape and Present-day Considerations 45

(3) Common International Standards for Anonymization Policy 46

Foreword

India stands at the forefront of the global race to harness the transformative power of artificial intelligence. With its vibrant digital economy, diverse talent pool, and rapidly expanding technology ecosystem, India is uniquely positioned to shape the future of AI, not just for its own citizens, but for the world. The choices made in India will influence how AI is developed, deployed, and governed across emerging and established markets alike.

Yet, as India accelerates its AI ambitions, it faces the same challenge as the rest of the international community: how to foster innovation while upholding the fundamental right to privacy. The Digital Personal Data Protection Act (DPDP) and the evolving AI policy landscape reflect India's commitment to both technological progress and robust data protection. Striking the right balance is not only a regulatory imperative, but a societal one—ensuring that the benefits of AI are realized without compromising individual trust or safety.

This report, a product of the Open Loop India Program, brings together insights from over 40 Indian companies working across the AI ecosystem to illuminate the path forward. It highlights the need for clear, pragmatic guidance that enables responsible AI development, while safeguarding the privacy and dignity of every individual through the sensible and efficient deployment of Privacy Enhancing Technologies. As India continues to lead in digital innovation, its approach to balancing AI advancement and data protection will serve as a model for the world.

The collaborative and evidence-based approach demonstrated by the Open Loop program should furthermore be considered a template for inclusive and robust lawmaking where complex technologies are being explored. I am grateful to Meta, TQH, The Dialogue, Tsaaro Consulting and the Data Security Council of India for leading this effort, as well as all the experts and organizations who contributed to this important work. Together, we can build an AI future that is innovative, inclusive, and anchored in respect for privacy.

Dr Sasmit Patra
Honorable Member of Parliament (Rajya Sabha)

About Open Loop

Meta's Open Loop is a global program that connects AI experts, policymakers, and companies to help develop effective and evidence-based policies for AI and other emerging technologies by gathering detailed feedback on new or existing policies, regulations, laws, or voluntary frameworks.

The aim is to improve the quality of guidance and regulation on emerging technologies, ensuring that they are understandable, feasible in practice, and likely to achieve their intended outcome.

Acknowledgements

The Open Loop India program was created and designed by Meta and implemented in collaboration with The Quantum Hub (TQH), The Dialogue, and Tsaaro Consulting. We would also like to acknowledge the support of the Data Security Council of India (DSCI) as an Advisory Partner.

We would like to specifically thank the following colleagues from TQH for their project management, research, event management and operational support on this program:

- Rohit Kumar | Founding Partner / Project Advisor
- Sumeysh Srivastava | Partner / Lead
- Ujval Mohan | Manager
- Akanksha Ghosh | Senior Analyst
- Salil Ahuja | Senior Analyst
- Aparna Joshi | Analyst
- Tithi Neogi | Analyst
- Supriya Shekher Azad | Analyst
- Raunaq Chandrashekar I Consultant

We would also like to extend our gratitude to our colleagues from The Dialogue for their research and operational support on this project:

- Kazim Rizvi | Founding Director
- Kamesh Shekar | Associate Director
- Kriti Singh | Associate Director - Programs and Operations
- Raunaq Sharma | Senior Research Associate

Additionally, we would like to thank our colleagues from Tsaaro Consulting for their operational support for this program:

- Akarsh Singh | Chief Privacy Officer & Founder
- Aditya Gautam | Director
- Aruna Vaz | CEO, Middle East
- Aditi Tiwari | Senior Consultant
- Greeshma MR | Senior Consultant
- Hiten Choudhary | Consultant

Special thanks goes to the Data Security Council for their advice on the design of this program:

- Deepa Ojha | Manager (Privacy and Policy)
- Shivangi Malhotra | Senior Associate – Policy & Privacy

We are also indebted to our group of experts from across the AI and competition ecosystems who shared their deep expertise and contributed to the development of the program's research approach, while noting that contribution does not equal endorsement of all points made in this report. Our thanks to:

- Alpan Raval | Wadhwani AI
- G.V. Anand Bhushan | Bhushan Rajaram
- Ankit Bose | Nasscom
- Bilal Mohamed | Future of Privacy Forum
- Chetan Arora | IIT Delhi, Yardi School of AI
- Kriti Trehan | Data & Co – Law & Policy Advisors
- Lalit Panda | Vidhi Centre for Legal Policy
- Nivedita Krishna | PACTA
- Pranesh Prakash | Anekaanta
- Shahana Chatterji | Shardul Amarchand Mangaldas & Co
- Tarunima Prabhakar | Centre for Responsible AI (CeRAI)

Finally, thank you also to our design partners at Craig Walker Design and Research, who helped us transform our data into a well-designed report.

Acknowledgements

Most importantly, we would like to thank the following organizations for their participation throughout the program — without their commitment and active involvement, this work would not have been possible:

Agami
apna.co
Appvin
bodyscans.in
Fuski.ai
Grexa.AI
Haqdarshak
IDInsight
Indus Action
Jhana.ai
M-League Limited
Mool AI
Next Billion AI
People+ai
Rocket Learning
Samagra
Sessionmate.ai
Spot Draft
Tagore AI
Tarisa
Vahan.AI
Wysa
Yoti
Youth Ki Awaaz
Zephara AI

A number of other companies participated in the research, but chose to remain unnamed.

Executive Summary

This report presents the findings and recommendations of The Open Loop India Program 2025, which convened a diverse cohort of organizations to examine regulatory questions surrounding Generative AI (GenAI) adoption and the Digital Personal Data Protection Act (DPDP Act), providing insights to inform policy development. The report documents how organizations are implementing data protection practices in AI development, providing insights that can inform India's approach to AI governance and data protection.

Key insights

1 AI's iterative nature requires flexible approaches to legal bases

Organizations report significant uncertainty (65%) when data collected for one purpose is later used for AI model training. Unlike conventional data processing where purposes are discrete and foreseeable, AI training purposes evolve as technology advances and new use cases arise. If a narrow interpretation of purpose is taken, this may inadvertently create uncertainty about whether broadly stated AI development purposes satisfy DPDP requirements for specific and informed consent, even when all activities serve an integrated objective. This uncertainty is further compounded by the use of web-scraped and internet-accessible data for model training, where nearly half of organizations (47%) report ambiguity around the scope of the DPDP Act's publicly available data exemption under Section 3(c)(ii), including what constitutes "publicly available" data and whether disclosure intent can be assumed at web scale.

2 Organizations are proactively anonymizing data but need regulatory clarity to scale these efforts

Companies apply varied techniques based on international standards (51%), sector norms (35%), and internal guidelines (47%). However, without a statutory definition of anonymized data under the DPDP Act that is clear, pragmatic and reasonably achievable, 59% seek clearer guidance to confidently scale their privacy-protective efforts.

3 Organizations are ready to adopt Privacy-Enhancing Technologies (PETs) but need regulatory recognition

While 67% use foundational privacy measures like anonymization and aggregation, only 5% use advanced PETs like differential privacy. The primary constraint is regulatory uncertainty (30%), amongst other factors. Organizations are prepared to invest in PET infrastructure once they have assurance that these investments will be recognized as valid compliance pathways.

Recommendations

Ⓐ

Issue an exemption for AI model training under Section 17(5)

Provide an interim exemption for AI model development activities from select DPDP provisions, particularly purpose limitation and consent requirements, while maintaining baseline protections. Organizations need explicit clarity on how existing lawful bases apply as data moves through iterative AI stages, including whether these activities constitute new purposes or extensions of the original one. This could cover initial training, fine-tuning, retraining, evaluation, and the ongoing improvement of AI systems. Several critical AI applications in areas such as public health, education, social protection, microcredit, and voice recognition depend on Indian personal data and cannot rely solely on synthetic data. In addition, the regulators could work with industry and AI experts to remove the restriction attached to the publicly available data exemption under Section 3(c)(ii) and to establish permanent and workable legal bases tailored to AI development.

Ⓑ

Issue clear guidance on anonymization for AI development

India has a strong opportunity to establish a statutory definition of anonymized data under the DPDP Act that is clear, pragmatic, and reasonably achievable. This would make anonymization a viable compliance pathway for organizations seeking to access and use data for responsible innovation. MeitY could issue guidance that recognizes anonymization as a spectrum and provides clarity on when data ceases to qualify as personal data under the DPDP Act. The guidance could reference internationally recognized approaches such as the EU GDPR's "means reasonably likely to be used" standard or the UK ICO's effective anonymization analysis, which considers technical and organizational measures, context of use, and intended purpose. The framework could provide clear guardrails that allow organizations to ensure individuals can no longer be identified, with obligations limited to reasonable efforts to prevent re-identification rather than strict liability. Organizations could be required to demonstrate, through testing and documentation, good faith attempts to achieve anonymization appropriate to their specific context and risk profile.

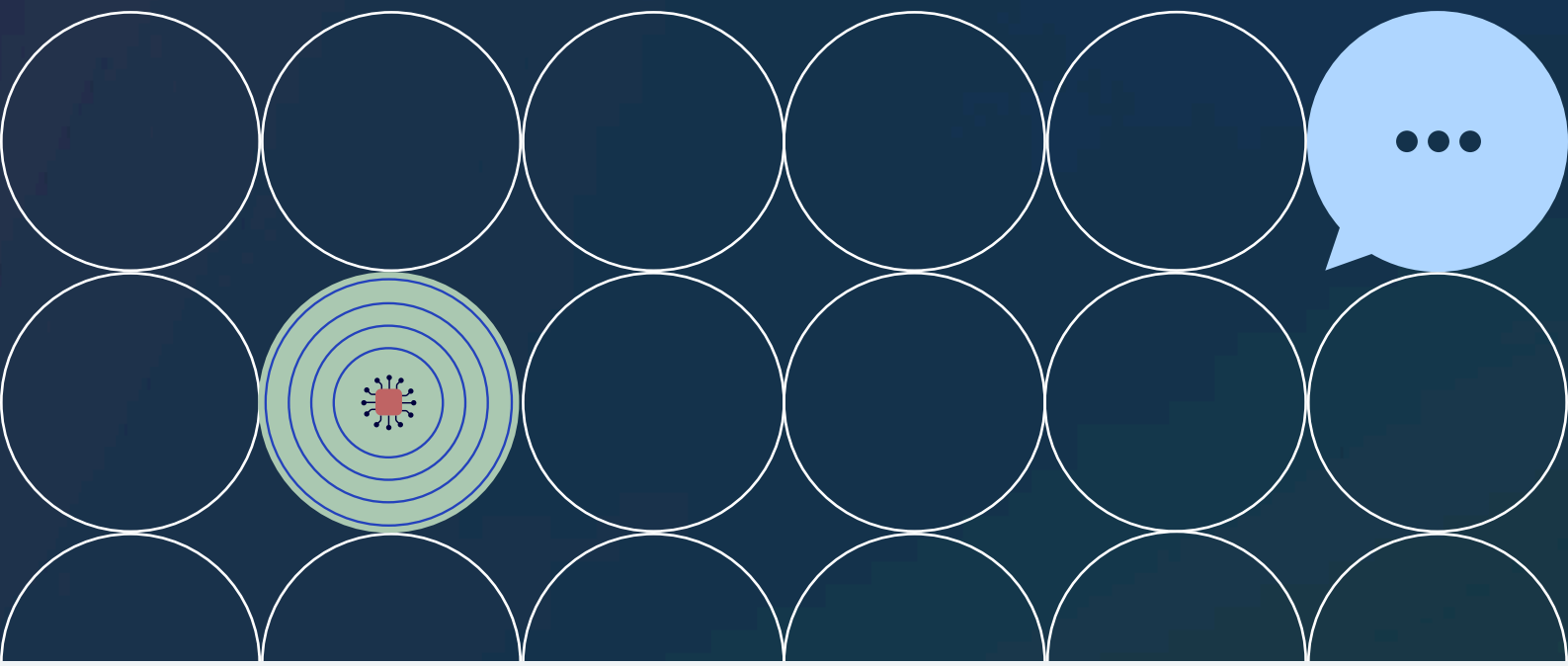
Ⓒ

Issue guidance on privacy-enhancing technologies (PETs) where appropriate

Issue guidance clarifying that properly implemented Privacy-Enhancing Technologies, including anonymization, differential privacy, federated learning, secure multi-party computation, and other privacy-preserving techniques, constitute reasonable security safeguards under Section 8(5) of the DPDP Act. The guidance could clarify two distinct pathways. Certain PETs may achieve full anonymization, allowing data to fall outside the scope of the DPDP Act, while others may operate as security safeguards where data remains within the framework. PETs should not be treated as universal solutions. Expectations should be outcome-focused, context-specific, and proportionate to organizational scale and risk. The guidance could also clarify that PET research qualifies as legitimate research under Section 17(2)(b) and provide for regulatory sandboxes to support testing and validation.

We thank all of the experts and companies who participated in the program and hope that the learnings and recommendations are useful to the Ministry of Electronics and Information Technology (MeitY), and other policymakers as they refine India's frameworks for both AI governance and data protection.

Introduction



Background

India is witnessing rapid transformation with the adoption of Generative Artificial Intelligence (GenAI) across critical sectors, such as banking, finance, healthcare, education, governance systems, and more. Its growing ubiquity has raised important questions around the access to and use of data required to train and improve GenAI models. The Digital Personal Data Protection Act, 2023 (DPDP Act) and the Digital Personal Data Protection Rules, 2025 (DPDP Rules) aim to govern these practices and have spurred discussions on how to adequately balance personal data protection and enable innovation. This presents a significant opportunity for stakeholders in India's digital economy to come together and offer considered inputs to inform policymakers and develop a robust framework that effectively harnesses GenAI's potential. The Open Loop India Program 2025 convened a diverse cohort of organizations innovating and deploying GenAI to better understand prevailing challenges relating to use of data and how the DPDP Act can be a vehicle for supporting and harnessing AI's catalytic potential in India.

In the past decade, policy discourse surrounding data use and protection has evolved considerably in India. Several iterations of a draft data protection legislation have been accompanied by reports and assessments from the Indian Parliament and government-constituted expert committees, which have included contributions from a diverse range of stakeholders. Most recently, the Principal Scientific Advisor to the Government of India (PSA) and the Ministry of Electronic and Information Technology (MeitY) released the India AI Governance Guidelines (2025) under the IndiaAI Mission. This follows the notification of the DPDP Rules, 2025, which operationalizes the provisions of the DPDP Act – making 2025 the most critical year for data regulation in recent years. These developments follow a significant proliferation of organizations leveraging GenAI.

Around the world, regulators are warming up to the idea of adapting traditional data protection frameworks to the specific contexts of AI training and development. The European Commission's Digital Omnibus Proposal acknowledges the need for AI-specific interpretations of personal data. The India AI Guidelines proposes regulatory sandboxes for industries with reasonable legal immunities, to enable them to innovate and build advanced AI technologies. It promotes privacy-by-design by embedding techno-legal regulation within system architecture and government-industry partnerships over voluntary frameworks. Yet, certain questions around data use and processing persist – specifically, how India's DPDP Act and Rules can be applied to the development and use of GenAI in a manner that both supports innovation and India's AI ambitions and protects citizens' personal data.

The DPDP Act's consent framework is designed for transactional data processing but does not provide AI-specific guidance on iterative data reuse across model training stages. Additionally, it does not describe PETs or their potential role, or establish a standard for determining whether an organization has anonymized data.[1]

Other regions are already testing different ways to think about personal data and the role of PETs. For instance, the European Court of Justice (CJEU) has determined that personal data is a "relative" concept (see Annex 1). Pseudonymized data is not personal data if it

1 The Annexures (Policy Background) elaborate on these in greater detail.

cannot reasonably be linked back to a person by the person holding the data and the potential means for linking, regardless of whether some other party may theoretically be able to do so.

At the center of today's debate: how can data protection frameworks accommodate the unique characteristics of AI development—including iterative workflows, evolving purposes, and novel technical approaches—while maintaining meaningful privacy protections?

This sits right where data protection and AI development intersect. Many now agree that GenAI works so differently that old data protection ideas don't always fit neatly.

India is dealing with the same tension. As its framework takes shape, these issues are coming to the fore. The DPDP Act does not clearly say when anonymized data stops being personal data. It also gives no AI-specific guidance on how to judge identifiability in model training.

This report offers a comprehensive look into these issues by spotlighting discussions with a diverse cohort of companies across India's AI ecosystem. A combination of mixed-methods surveys, multi-stakeholder consultations, and interviews highlight India's opportunity to devise AI-specific regulations that can enable its burgeoning AI ecosystem in a manner that supports both innovation and user privacy.

Issues/Research Questions

The Open Loop India Program was a consultative exercise that focused on three core questions:

How do organizations collect, manage, and safeguard personal (and non-personal) data in AI workflows?

How and when do they utilize PETs (including anonymization)?

- What operational challenges and compliance risks exist under the current regulatory framework?

What type of regulations and related measures can best enable responsible and effective AI innovation across India?

- Which facets of the DPDP Act and Rules require further adaptation to AI, which can spur greater innovation and mitigate compliance risks?

Approach

The Program sought to understand how Indian organizations view these issues, such as data privacy, anonymization, and compliance under the DPDP Act and Rules as they develop their AI products. To ensure diversity and representation, the program included over 44 organizations of varying sizes across 13 sectors. Cohort members answered a structured online survey that captured qualitative and quantitative trends. It comprised four sections:

1. **Profiles**
   The nature and size of organizations and their data governance and use practices, including data sources, handling, and protection mechanisms.

2. **Compliance**
   Compliance readiness and experience under the DPDP framework.

3. **Data Protection Practices**
   The knowledge and use of PETs and anonymization practices.

4. **Challenges**
   Ongoing challenges faced and additional regulatory guidance required.

The survey included multiple-choice, ranking, and open-ended questions to capture broad trends, specific practices, and nuanced perspectives. It was accompanied by semi-structured interviews conducted with 14 cohort companies. The report incorporates secondary research and analysis of the Indian policy landscape and comparable global approaches, including consultations with technical, legal, and policy experts.

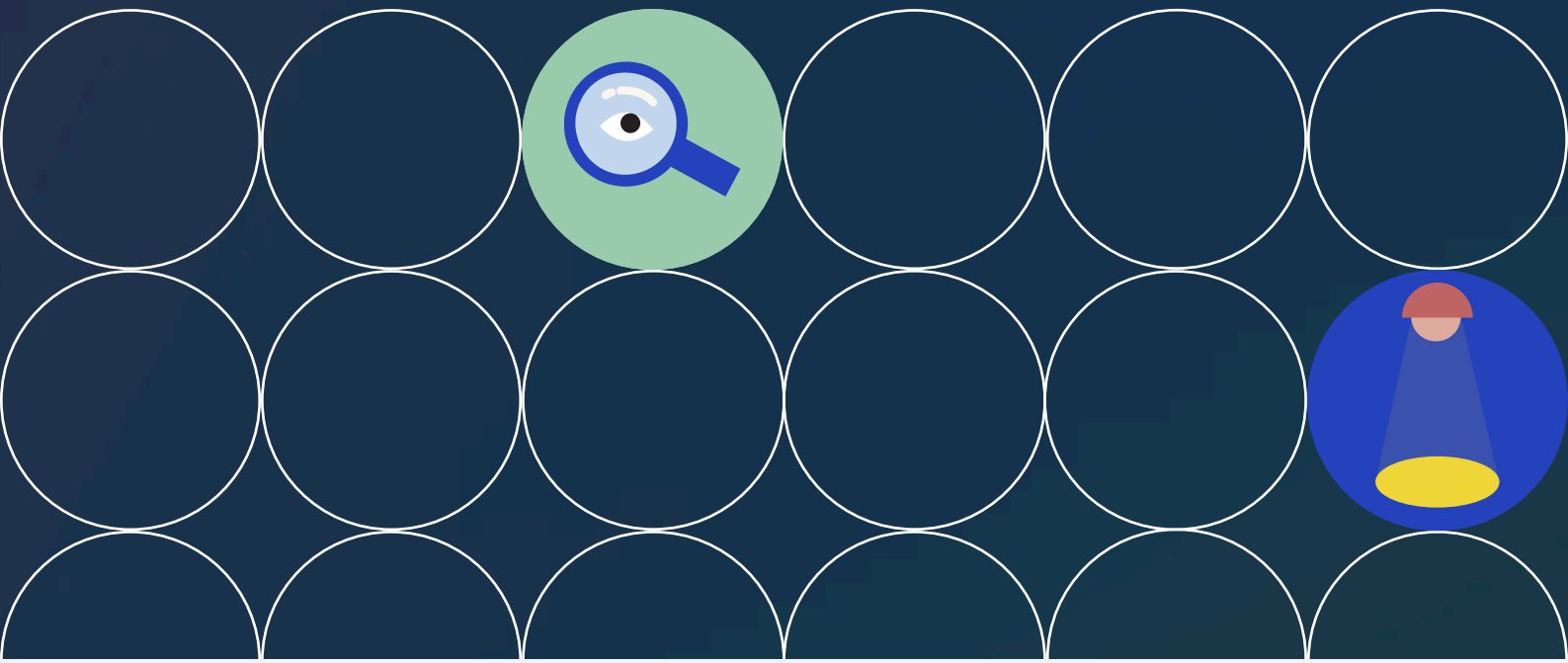
A detailed explanation of the methodology can be found in Annex 2.

Report structure

The report is divided into four chapters and three annexes:

Chapter 1: Introduction	This section introduces the India Open Loop program 2025 and presents a brief overview of India's policy landscape, alongside some global and local developments in how data protection frameworks have gradually adapted to AI.
Chapter 2: Findings and Observations	This section presents a detailed analysis of the survey, interviews, and expert consultations, providing a comprehensive look into how organizations perceive India's AI regulatory framework and its impact on the ecosystem.
Chapter 3: Recommendations	This section draws on insights and observations from the cohort to offer actionable recommendations that policymakers can integrate into India's data regulation framework in a manner that supports AI innovation while securing privacy.
Chapter 4: Conclusions	The final section examines the transformative potential of the recommendations and highlights considerations that can be explored by Indian policymakers and organizations.
Annex 1: Glossary (Key Concepts)	Definitions and explanations of the various technical concepts and terms used in this report.
Annex 2: Research Methodology	A detailed description of the mixed methods research and analyses used to create this report and generate its findings, observations, and recommendations.
Annex 3: Policy Background	An examination of key DPDP Act provisions and related governmental regulations and guidelines governing data use and AI.

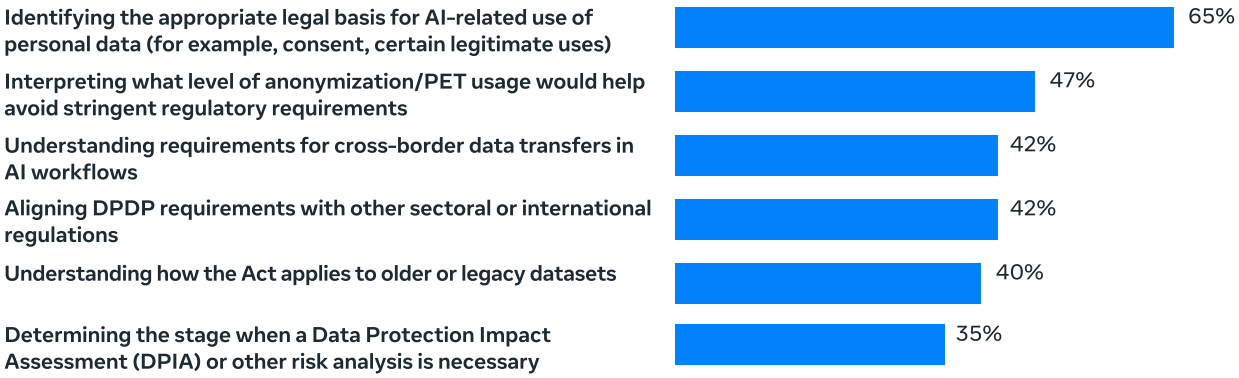
Findings & observations



1 AI's iterative nature requires flexible approaches to legal bases

1.1 What companies are already doing

AI development is inherently iterative as models are continuously refined through training, evaluation, fine-tuning, and deployment. Data collected for one purpose is often reused across these stages, creating uncertainty about whether integrated AI development activities might be interpreted as multiple separate purposes. The Digital Personal Data Protection Act, 2023 provides several pathways for lawful processing: explicit consent, 'legitimate uses' outlined in Section 7, publicly available data under Section 3(c)(ii), or other exemptions under Section 17. However, organizations report uncertainty (65%) about which pathways apply as data moves through iterative AI workflows.



Note: Each respondent selected more than one option

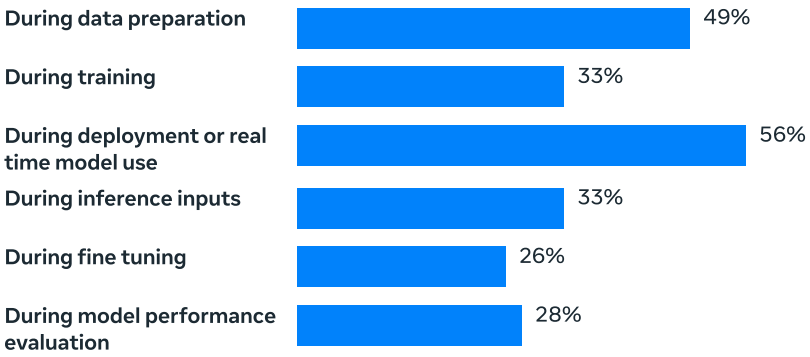
The issue is that as data moves through the AI lifecycle, it is used and repurposed in ways that are intrinsic to the iterative nature of AI development. While consent can be obtained at the point of collection, static consent mechanisms are ill-suited to exhaustively specify every future model refinement or system improvement at that stage (Mohamed, 2024). If “purpose” is interpreted too narrowly, these predictable stages of development risk being treated as separate purposes, making it difficult to meaningfully anchor consent to a single, appropriately defined purpose.

Within this context, we see that when personal data processing for AI has a single, defined purpose, existing practices under the DPDP Act function as intended. Most cohort companies have established lawful bases for these baseline uses, and many disclose anticipated AI-related processing upfront. Problems arise when data moves into the iterative workflows that define AI development. As one company put it, "The primary uncertainty is in reconciling the AI lifecycle with the DPDPA's strict requirements…Purpose limitation is hard to apply when AI purposes evolve from transactional to developmental, like moving from processing a customer request to using that interaction for model training." Thus, for AI model development specifically, this underscores the need for purpose to be interpreted at an appropriate level of generality.

Accordingly, while companies can and do seek consent upfront for AI development and product improvement, uncertainty persists where consent and purpose limitation are interpreted in an overly granular or stage-specific manner, rather than at a level of generality that reflects the iterative nature of AI development.

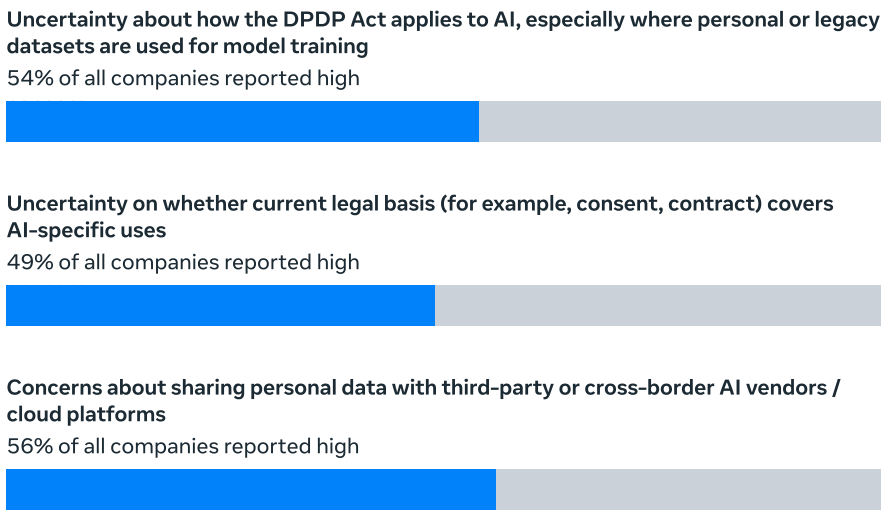
1.2 What the data tells us about how AI workflows interact with lawful bases

AI systems in the cohort routinely process personal data across pre and post deployment stages of their lifecycle:



AI development involves integrated activities in which the same data may be processed at collection, during training, fine-tuning, and deployment. Traditional data processing contexts often involve discrete and immediate purposes, such as “process payment” or “send notification.” Some developers may conceptualize processing in terms of these granular, immediate purposes. However, AI development differs fundamentally. The purpose of “develop and improve AI systems” is inherently iterative and long term, with specific technical implementations evolving over time. If a narrow interpretation of purpose is taken, this may inadvertently create uncertainty about whether broadly stated AI development purposes satisfy DPDP requirements for specific and informed consent, even when all activities serve an integrated objective.

Survey responses confirm that uncertainty concentrates not around baseline compliance, but around whether existing lawful bases continue to apply as data is reused across AI lifecycle stages. When cohort companies were asked to rank risks associated with personal data use in their AI operations (1 = no concern; 5 = extremely high concern), the three highest-scoring concerns (avg. score of 4) were:



1: AI's iterative nature requires flexible approaches to legal bases

For each of these, over half of respondents reported being very or extremely concerned. Only 5% reported no concern around how the DPDP Act applies to AI use cases, indicating near-universal requirement for clarity.

These patterns show that the issue is not lack of compliance intent, but uncertainty about how existing lawful bases apply as data moves through AI-specific workflows. As one organization noted, “We collect data to deliver a service. Later, we de-identify this data as it is useful for improving the model. But is our level of de-identification sufficient, or are we still covered by the original consent obtained for personal data?” Where purpose is interpreted too granularly, organisations may be forced to reassess their legal basis at each stage of model development, even when these activities are undertaken in furtherance of the same broadly defined purpose and the nature of the processing remains consistent.

Additionally, nearly half of the companies surveyed (47%) highlighted concerns regarding legal uncertainty around the use of web-scraped data, particularly where such data may include personal or sensitive information. Organizations developing AI systems seek to use data accessible on the internet for model training but face uncertainty about the scope of Section 3(c)(ii), which exempts personal data “made or caused to be made publicly available by the Data Principal.” Key questions include what constitutes “publicly available” data in practice, and whether data must be voluntarily disclosed by individuals in a specific manner. This ambiguity is particularly acute for organizations training models on web-scale data, where verifying the disclosure intent for each data element is impractical.

1.3 What challenges companies are facing

Uncertainty in applying purpose limitation across iterative AI workflows uncertainty around:
As highlighted above, consent under the DPDP Act must be specific, informed, and tied to a defined purpose. In AI development, that purpose is often articulated at an appropriate level of generality, even though the specific stages and technical processes through which it is realized may evolve over time. Alternatively, for the purpose of AI model development, companies may use de-identified or anonymized data, allowing them to step outside the scope of consent frameworks for certain processing stages. However, while this approach may reduce privacy risks, there is currently limited clarity on when anonymized data can be confidently treated as no longer constituting personal data under the DPDP Act, making it difficult for organisations to rely on anonymization as a clear compliance boundary. Organizations seek clarity on anonymization standards particularly suited to AI training contexts. One approach is to recognize that when unstructured data used for AI training is de-linked from account identifiers, personally identifiable information, or any specific identifiable person, it may achieve effective anonymization. Even where the unstructured content itself may contain personal information, once it is separated from identifiers that would allow linkage to specific individuals, the re-identification risk may be reduced to a level where the data should no longer be treated as 'personal data' under the DPDP Act. This recognizes that AI training on unstructured, de-linked data does not pose re-identification risks comparable to traditional personal data processing where individual records remain intact and linkable.

Lack of clarity leads to conservative data strategies and design trade-offs:
In response to this uncertainty, organizations adopt risk-minimizing data strategies such as early hashing or redaction, minimal retention, heavy use of synthetic or sanitized datasets, and tightly restricted processing environments. These privacy-first design choices meaningfully reduce exposure, but they also restrict personalization and limit model performance. One company explained:

> "Sometimes we avoid storing personal data altogether and rely on synthetic generation - 20–30% real and the rest created by an LLM. But nothing beats user-generated data, and edge cases don't appear clearly in synthetic datasets."

These decisions meaningfully reduce exposure, but they also restrict personalization, slow model improvement, reduce coverage of real-world scenarios, and introduce accuracy or latency constraints. Another cohort company shared:

> "Our work is deeply contextual, socio-economic realities, gender norms, caste dynamics. Synthetic data can't replicate that, and it can even bake in new biases. You simply can't replace real Indian data for many AI tasks."

1.4 What this means for the ecosystem:

> Where purpose limitation and consent requirements are interpreted narrowly or applied in a stage-specific manner, organisations may default to overly cautious compliance approaches, even when processing remains aligned with a broadly defined, lawful purpose such as product or AI system development.

> Overly conservative data strategies minimize exposure to risk but limit experimentation, personalization, post-deployment improvement, and the use of real-world datasets that reflect India's diverse socio-economic contexts.

This highlights three complementary needs:

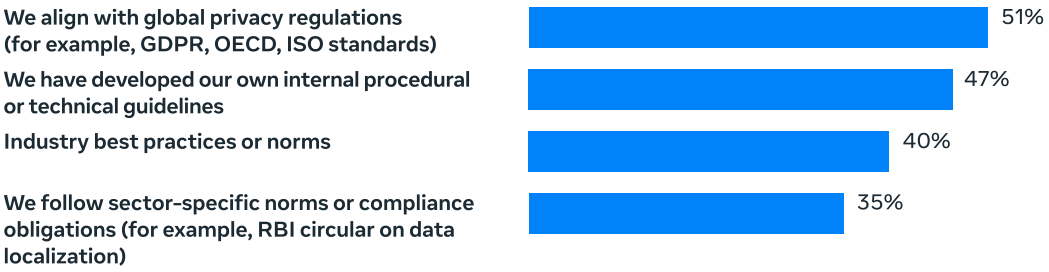
1. Explicit clarity on how existing lawful bases apply as data moves through iterative AI stages, including whether model training, fine-tuning, or performance evaluation constitute new purposes or extensions of the original one.
2. Clarity on the scope of the publicly available data exemption under Section 3(c)(ii), particularly in relation to data accessible on the internet that organizations seek to use for model training.
3. Where anonymization is technically feasible and appropriate, clear standards that allow organizations to confidently determine when data has been effectively anonymized. Without clarity across all three dimensions, namely lawful bases for iterative workflows, publicly available data, and anonymization standards, organizations will continue to face structural uncertainty that constrains responsible innovation.

2 Clear anonymization standards are critical for activities across the AI lifecycle

2.1 What companies are already doing

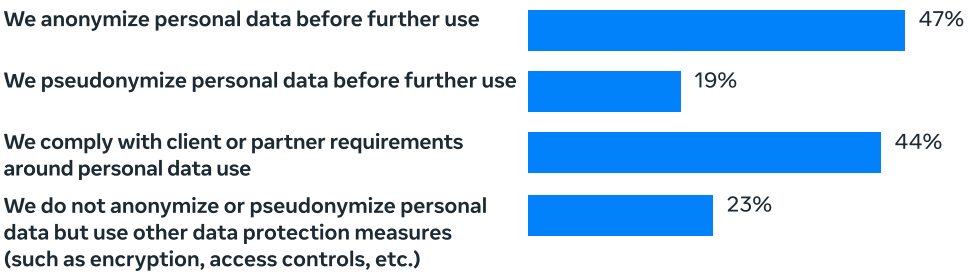
As identified in Insight 1, purpose limitation creates friction across AI's multi-stage development workflows. One potential compliance pathway (where technically feasible and appropriate for specific use cases) is anonymization. Whether anonymization is suitable depends on factors including the AI activity, the nature of the data, model performance requirements, and organizational context. Recognizing this, companies across the ecosystem are proactively de-identifying personal data to enable responsible AI development. However, anonymization practices vary significantly because there is no prescribed statutory benchmark for what constitutes "effective anonymization" under the DPDP Act.

Organizations currently align their de-identification choices with a range of reference points, including:



Note: Each respondent selected more than one option

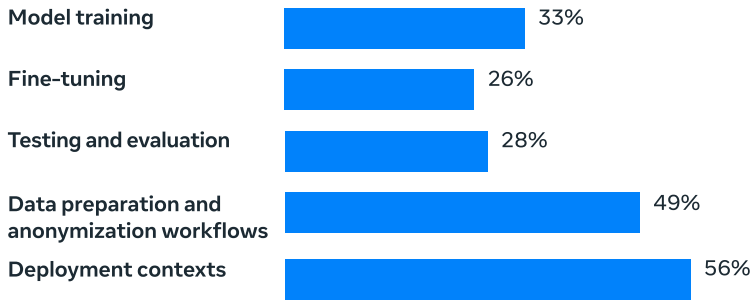
Some companies in our cohort have reported that they follow international standards such as the EU's "reasonably likely" test (derived from Recital 26 of the EU GDPR), while others adopt techniques calibrated to internal risk perceptions and available resources. Organizations employ multiple approaches across a spectrum:



Note: Each respondent selected more than one option

2.2 What data tells us

Survey responses indicate that organizations handle personal data across the AI lifecycle, including:



Nearly half of the respondents said that they proactively rely on anonymization as a technique for preserving privacy while handling personal data for AI purposes. However, using anonymization also leads to certain trade-offs. For instance, 35% of respondents reported that stringent anonymization leads to reduced model performance and accuracy, and 37% flagged that it leads to increased costs.

Across the survey, one insight that stood out was that companies asked for more clarity to understand if their anonymization practices were legally sufficient from a compliance standpoint. 56% of all companies participating in the survey stated that regulatory clarity particularly in the form of applicability to certain use cases, and guidance on technique and standards would be extremely beneficial to foster adoption of PETs or anonymization for AI. Companies expressed that, without a clear anonymization benchmark, they hesitate to proceed confidently with these activities, even when technically capable of doing so responsibly.

2.3 What challenges companies are facing

Despite this proactive approach to anonymization, a fundamental challenge exists: Companies apply varied techniques based on their interpretations of what is de-identified or anonymized data, leading to divergent practices and an ecosystem whose understandings are not harmonized.

- **Companies implement measures based on their capabilities, interpretation, and validate their approaches via internal testing**, as they do not have a prescribed statutory test to serve as a recognized benchmark. The survey showed that 67% use internal testing, 49% use legal/ compliance team reviews, and 26% rely on vendor assurances. The diversity is a result of anonymization techniques varying across a spectrum.
- **Organizations don't have confidence in their current privacy safeguards**, largely driven by lack of adequacy benchmarks. When asked to rate how confident they were that their current privacy safeguards effectively prevent re-identification or unintended reproduction of personal data [on a scale of 1 (not confident) to 5 (extremely confident)], 61% of organisations rated themselves at 3 or below. This shows that without express regulatory benchmarks, it is difficult to assess residual re-identification risks in data types like text, audio, or images with full certainty.

Another company shared, “We can’t be as good with personalization if we don’t have enough identifiable data. We can mask it for humans, but the system sometimes still needs it. There’s always a trade-off. For instance, child speech is very different from adult speech. If we can’t store some of it, we can’t fine-tune models properly. But storing it creates new risks. So, there’s a constant balance.”

- **The type of clarity needed can vary by sector:** healthtech organizations require frameworks for using de-identified health data for model development, while Banking, Financial Services, and Insurance (BFSI) companies need AI-specific data governance controls, and IT firms seek harmonization with global frameworks like GDPR. Cohort companies across sectors, such as healthtech, fintech, edtech, and others, expressed this uncertainty clearly, stating:
  - “Our clinical data is de-identified according to international standards, but we don’t know if that’s sufficient under DPDP. The Act doesn’t specify technical requirements, so we’ve paused our diagnostic AI initiatives.”
  - “We are not against compliance. The problem is we don’t know what is considered ‘enough’ – are we supposed to anonymize, pseudonymize, or encrypt? We do not have a reference point.”
  - “We need more clear guidance on how to, what to and to what extent we should anonymize. There should be a definitive proper guideline based on use cases.”
  - “We pseudonymize data, but DPDP does not define what counts as ‘effective anonymization’ and without it, we are not certain if we have exited the regulatory scope.”

Consequently, companies are not certain as to when personal data, when de-identified using one or more techniques, ceases to be “personal” as per the DPDP Act.

2.4 What this means for the ecosystem

The absence of a consistent anonymization benchmark creates long-term friction for AI capability development. This friction runs across the entire AI lifecycle, but it shows up differently at different stages. It is most acute during initial model training, which involves large volumes of data processed over long timeframes, with purposes that often cannot be fully defined at the point of collection. Organizations consistently describe this as their biggest compliance uncertainty. In later stages such as fine-tuning, testing, and deployment, the challenge looks different. Here, clearer anonymization standards could offer workable compliance pathways where anonymization is technically feasible.

Organizations are concerned that without clear benchmarks today, future regulatory interpretations may differ from current practices, potentially requiring re-engineering of existing approaches and creating operational and financial uncertainty.

Clear, proportionate anonymization standards including a recognized statutory test (such as the ICO’s effective anonymization analysis or EU “reasonably likely” threshold) would help harmonize practices across sectors and create predictable, auditable compliance pathways. This clarity is also likely to enable confident expansion of AI-enabled products and collaborations.

Taken together, this points to a dual requirement. First, the purpose-limitation friction in training needs to be addressed through targeted regulatory mechanisms, acknowledging that anonymization may not always be appropriate for training objectives. Second, clearer anonymization guidance is needed for AI activities where anonymization is both feasible and appropriate.

3 High PETs awareness but limited adoption due to barriers

3.1 What companies are already doing

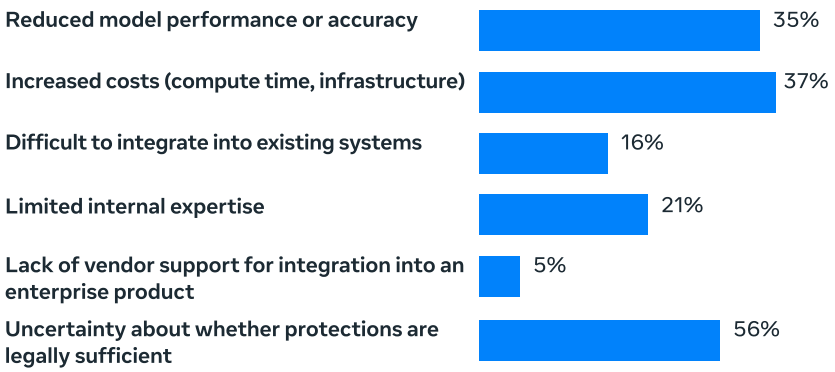
Organizations across India's AI ecosystem demonstrate strong awareness of Privacy-Enhancing Technologies (PETs) as tools for protecting privacy while enabling data utility for AI development. Companies recognize that PETs, ranging from foundational techniques like anonymization and aggregation to advanced methods like differential privacy and federated learning, can help achieve effectively anonymized data in appropriate circumstances, supporting the principle that anonymization exists on a spectrum rather than as a binary state.

Survey results show that most organizations use foundational privacy measures including anonymization (67%) and aggregation techniques (67%), and 42% report being familiar with PETs and applying them actively in their AI workflows. Privacy practices in AI are shaped by both internal data management practices and external pressures. Organizations balance their own privacy principles (53%) against regulatory requirements from global laws (63%) and the DPDP Act (56%), as well as client expectations (58%).

3.2 What challenges companies are facing in implementing PETs

Despite this awareness, adoption of advanced PETs remains limited. Only 5% of organizations currently employ sophisticated techniques like differential privacy. The barriers to broader adoption are structural and regulatory rather than a lack of motivation or technical understanding.

When asked about obstacles to wider PET deployment, respondents consistently cited:



Note: Each respondent selected more than one option

These challenges have led many organizations to adopt a phased approach, experimenting with basic safeguards while postponing advanced techniques until explicit regulatory guidance, affordable tools, and sector benchmarks become available.

3: High PETs awareness but limited adoption due to barriers

The regulatory uncertainty creates a practical dilemma for organizations. A cohort company shared that it was still evaluating the use of PETs such as synthetic data and was awaiting notification of the DPDP Rules for performing Data Protection Impact Assessments (they were interviewed before the rules were notified). They were waiting for express regulatory guidance on PETs before adopting more advanced tools, saying they were “unsure of effectiveness on personal data and unwilling to risk experimentation.” This shows that, without regulatory clarity on what counts as PETs and their proper use cases, excessive risk aversion sets in and blocks the experimentation needed for high-value AI applications.

A cohort company also pointed out that PET adoption often involves trade-offs. “Our applications serve low-literacy users in semi-urban and rural areas. The heavier the PET layer, like encryption, consent tokens, and de-identification, the harder it becomes to maintain usability and human assistance. It’s a constant trade-off.” This shows that the suitability of specific PETs depends heavily on context.

3.3 What this means for the ecosystem, and its significance

The data points to a maturity gap. While 40% of respondents said they are familiar with PETs and use them actively, only 5% use advanced techniques like differential privacy. High costs, limited expertise, and the lack of clear regulatory benchmarks together discourage experimentation with advanced PETs such as differential privacy or federated learning.

Right now, there is no clear regulatory guidance on when PETs count as “reasonable security safeguards” under Section 8 of the DPDP Act. Organizations need clearer direction on when they can use personal data to test and validate their risk controls. Without clarity on whether PETs meet compliance needs and how they should be used, many hesitate to invest in PET infrastructure whose compliance value is still uncertain.

Additional findings

Beyond the three challenges outlined above, the research revealed observations in additional areas that can inform broader AI governance discourse:

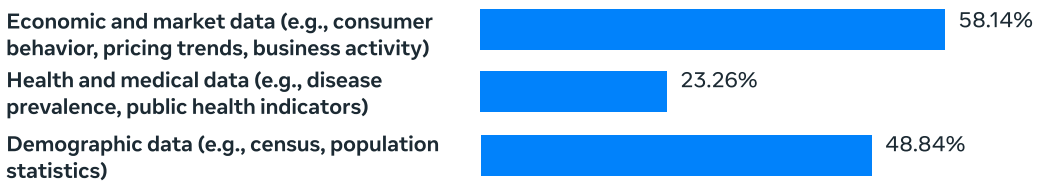
4 Conventional privacy compliance models may not translate well to all contexts

Organizations serving marginalized or low-connectivity/remotely-located populations operate in assisted models where community workers help users navigate services. These organizations have adapted consent mechanisms to context OTP-assisted (One Time Password) verbal consent, pictorial notices, facilitated explanations etc. but need clarity about whether these approaches satisfy DPDP requirements. Standard digital consent flows designed for informed, self-service users can disrupt trust relationships with low-literacy populations. One company reported that repeated consent requests at every stage inadvertently reduces trust. While community workers build trust with users and secure consent in the first instance, they eventually introduce hesitance by repeatedly prompting them with consent requests, which makes users apprehensive/suspicious. Organizations emphasized the need for context-specific guidance on valid consent mechanisms for assisted digital environments and clarity on when strong technical safeguards may satisfy compliance requirements without repeated consent prompts that alienate vulnerable populations.

5 High-quality, domain-specific Indian datasets remain limited

5.1 What the data tells us about companies using public datasets

While more than half of respondents use Indian public datasets for Indian users, critical gaps exist between available data and data required for India-specific AI development. Respondents most frequently cited:



Note: Each respondent selected more than one option

Other top datasets cited as essential for AI innovation were multilingual conversational and profanity corpora, structured legal data (case law, e-courts, legislation), and granular public-sector datasets (scheme data, location data).

5 High-quality, domain-specific Indian datasets remain limited

Across sectors, primary concerns centered on data quality and accessibility. Health-tech organizations noted gaps in usable de-identified health data, while legal-tech and civic-tech organizations highlighted the need for machine-readable datasets from legislatures and judicial bodies. About 26% cited limited technical expertise to utilize data, while interview responses consistently yielded issues with public data infrastructure, including uniformity in datasets.

5.2 Relevance of public datasets for the ecosystem

High quality public datasets that reflect linguistic and regional diversity are crucial to enable personalized offerings of AI development for Indian users, a claim strongly substantiated by company respondents in the survey and interviews. However, public datasets are not easily accessible - and our findings indicate that the issue lies in public data infrastructure rather than limited organizational capabilities. This presents an opportunity to build public, aggregated, sector-specific datasets for AI development serving socio-economic development and governance outcomes.

5.3 Challenges and considerations for accessing public datasets

Innovators and the tech ecosystem would benefit from clear regulatory positioning on access to public datasets for AI usage. While policy inputs have emphasized AI's value as a potentially transformative public good (India's National Strategy on AI is an early policy input that stresses this), MeitY's recent stance on AI copyright licensing that public benefit cannot be a qualifying criterion for innovation or grant legal immunity to innovators (Bansal, 2025) - even at the cost of scaling these technologies - builds confusion for innovators who wish to scale their public-interest technologies. Policy inputs from public bodies, such as the latest report by the Principal Scientific Adviser (PSA) (2025) are already advising for a democratic approach to foundational AI resources, to enable responsible AI innovation. This democratic approach needs to be uniformly sustained and promoted alongside existing initiatives (such as MeitY's AIKosha) to improve public data access, to build greater confidence among AI innovators.

Government efforts to address these challenges are underway. IndiaAI Kosh (launched in 2025) serves as a national repository across 20 sectors and has onboarded 5,722 datasets and 251 models. Bhashini provides language datasets and models for Indian languages, hosting over 350 language AI models. State-level initiatives like Telangana Data Exchange (TGDeX) demonstrate federated models for dataset sharing. The National Data & Analytics Platform (NDAP) consolidates government datasets in machine-readable formats.

These initiatives address some accessibility challenges, but significant work remains to close critical data gaps, improve quality, and expand coverage in sectors like economic data, healthcare, and legal datasets where respondents report the most acute needs.

Recommendations

Recommendations

The following recommendations address the three critical challenges identified through this research: the need for anonymization guidance, promoting PET adoption through express guidelines, and the friction to AI development workflows when these features are not present explicitly in regulation.

RECOMMENDATION

A Issue exemption for AI model training under Section 17(5)

DETAILS

Organizations report significant uncertainty (65%) when data collected for one purpose is later used for AI model training. Unlike conventional data processing, where purposes are typically discrete and reasonably foreseeable, AI training purposes are often difficult to fully specify at the point of data collection because model development and improvement needs evolve as technology advances. This creates acute regulatory friction. Organizations need explicit clarity on how existing lawful bases apply as data moves through these iterative AI stages—specifically, whether model training, fine-tuning, or performance evaluation constitute new purposes or extensions of the original collection purpose. Consent frameworks struggle to accommodate the iterative and continuous nature of model training when purposes evolve beyond what was specified at collection, and existing categories of legitimate use and exemptions under the DPDP framework do not explicitly account for AI development activities.

India's AI strategy aims to address India-centric challenges through development of India-specific datasets and models, as articulated in the India AI Mission launched in 2024. As discussed earlier in this report, critical AI applications in public health, education, social protection, micro-credit, and voice recognition depend on access to Indian personal data and cannot rely solely on synthetic or anonymized data without undermining model effectiveness. Absent regulatory clarity, this uncertainty risks delaying AI development in precisely the areas that India is keen to prioritize.

An interim mechanism is therefore required to provide regulatory certainty for innovation in the near term, while allowing the ecosystem time to mature. During this period, the government can work with industry and research institutions to develop clear anonymization guidance (Recommendation B), encourage adoption of Privacy-Enhancing Technologies (Recommendation C), expand the scope and remove restrictions attached to the publicly available data exemption, and assess where personal data remains necessary for model training and how it should be appropriately handled. This approach can enable responsible experimentation and learning, informing the development of permanent frameworks that balance innovation with privacy protection.

A critical area requiring clarity is the scope of Section 3(c)(ii), which exempts personal data "made or caused to be made publicly available by the Data Principal." The current requirement that public

data must be "voluntarily disclosed" by the user creates significant uncertainty and operational friction for organizations seeking to use publicly accessible internet data for AI model training. India could adopt approaches similar to Singapore, which defines publicly available data as information "generally available to the public," (Singapore PDPA, 2012) or US state laws like Utah, which recognize data a controller "reasonably believes" has been "lawfully made available to the general public." (Utah Consumer Privacy Act, 2022) This approach acknowledges that it is impractical to individually verify each data element on the internet, while maintaining protections against misuse of truly private data. The clarification could confirm that data publicly accessible on lawfully operated websites, not restricted to specific audiences, and not obtained through circumvention of technical barriers, qualifies for the exemption. This will align India's framework with global best practices and enable responsible use of publicly available data for AI development.

Section 17(5) of the DPDP Act provides a statutory mechanism for this transition: "The Central Government may, before expiry of five years from the date of commencement of this Act, by notification, declare that any provision of this Act shall not apply to such Data Fiduciary or classes of Data Fiduciaries for such period as may be specified in the notification."

Using this provision, the Central Government could issue a notification under Section 17(5) exempting Data Fiduciaries engaged in AI model training from select DPDP provisions (particularly purpose limitation and consent requirements for using collected data for model training activities), until a regulatory framework is specifically created for this purpose.

The AI Governance Guidelines signal the government's intention to adopt a phased regulatory approach: relying on existing laws and voluntary measures in the short to medium term and developing comprehensive frameworks for the long term. The exemption proposed here aligns with that adaptive strategy, enabling innovation in the interim while the government develops permanent lawful bases under the DPDP framework specifically tailored to AI model training. These could include new legitimate use grounds, expanded research exemptions, or AI-specific provisions through amendments.

The exemption can be narrowly defined and targeted for core model development activities:

1. Initial training of foundation models, domain-specific models, or fine-tuned models using datasets
2. Model evaluation, validation, and performance testing
3. Research and development of model architectures and training methodologies

Such a structure for the exemption can address the purpose limitation friction, subject to the application of baseline protections such as data security obligations, data principal rights, and enhanced requirements for any other categories of data requiring heightened protection. Organizations can use anonymization pathways (Recommendation B) supported by Privacy-Enhancing Technologies (Recommendation C) where technically feasible and appropriate for their training objectives. Where anonymization is not suitable for specific training needs, this exemption provides a compliant pathway. MeitY can use the learnings from this period to assess whether the anonymization guidance and PET frameworks developed under

Recommendations B and C provide viable alternative compliance pathways, whether clarification on the publicly available data exemption addresses a significant portion of industry needs, and what permanent lawful bases are needed for cases where personal data remains necessary for model training, and whether any aspects of the temporary exemption should be converted to permanent provisions.

This approach recognizes that while comprehensive regulation is the long-term goal, immediate clarity is necessary for India's AI ecosystem to remain competitive. The exemption provides that immediate clarity for a limited, well-defined activity while the government develops broader frameworks through collaborative assessment of technical capabilities, privacy safeguards, and innovation needs.

RECOMMENDATION

B Issue clear guidance on anonymization for AI development

DETAILS

Organizations seek clearer guidance on anonymization requirements (61%) and regulatory certainty (65%) to advance their AI development. Without a statutory definition of what constitutes "effectively anonymized" or "effectively de-identified" data, organizations employ varied techniques—some aligned with international standards like ISO/IEC 20889 (ISO/IEC 20889:2018(en), Privacy enhancing data de-identification terminology and classification of techniques), others based on internal judgment—without certainty about whether their approaches satisfy DPDP requirements. Nearly half (49%) reported that clear anonymization standards would reduce perceived regulatory risk and unlock currently-stalled AI projects.

The India AI Governance Guidelines acknowledge this ambiguity and call for techno-legal approaches that embed compliance into system design. The Guidelines position the Technology & Policy Expert Committee (TPEC) to collaborate with standards bodies and industry on sector-specific frameworks, while the AI Safety Institute (AISI) develops India-specific testing methodologies. This recommendation operationalizes institutional architecture specifically for anonymization benchmarks.

MeitY could issue clear guidance on anonymization for India. Given the technical complexity and need for diverse expertise, this could be done in consultation with industry, civil society, and academia. The guidance could adopt a risk-based approach similar to internationally recognized standards, establishing that data ceases to be "personal data" under the DPDP Act when re-identification risk is reduced to a level where it is no longer reasonably likely given the technology, resources, and context available at the time of assessment. Specifically, the guidance could align with internationally recognized standards such as the EU GDPR's 'means reasonably likely to be used' formulation or the UK ICO's effective anonymization analysis, which assesses multiple factors including the motivated intruder test alongside technical and organizational measures, context of use, and intended purpose. Additionally, it could recognize anonymization as a spectrum rather than a binary state, with adequacy assessed based on whether re-identification risk is acceptably low for the intended use, considering available technology and context. The framework could provide clear guardrails within which organizations can adopt measures to ensure individuals can no longer be identified through the de-identified or anonymized data. The onus on organizations should be limited to reasonable efforts to prevent re-identification, rather than strict liability or a near-zero possibility of re-identification. Organizations could be required to demonstrate through testing and documentation that they have made good faith attempts to achieve anonymization appropriate to their specific context and risk profile.

The anonymization guidance should enable proportionate application based on the intended use and context of data processing. Organizations processing data for internal analytics with limited re-identification vectors face different considerations than those releasing datasets publicly or processing highly sensitive information. The framework should allow organizations to calibrate their anonymization approaches to actual risk while providing clear benchmarks for common scenarios.

Additionally, formal definitions distinguishing anonymization, de-identification, and pseudonymization should be issued. These definitions should be accompanied by practical worked examples from sectors where data sensitivity and AI applications intersect most visibly: healthcare, financial services, and education. The examples should address common scenarios organizations face:

- Using historical datasets for AI model training
- Sharing data with vendors or across organizational units for collaborative AI development
- Handling legacy data collected before the DPDP Act
- Processing different data modalities (text, voice, images) that each present distinct re-identification challenges

Sectoral regulators could coordinate to develop sector-specific guidance while maintaining a consistent underlying framework. India-specific protocols could be developed and this could include creating benchmark datasets, attack simulation methodologies that test resilience against re-identification attempts, and documentation frameworks that enable organizations to demonstrate compliance.

RECOMMENDATION

C Issue guidance on privacy-enhancing technologies (PETs) where appropriate

DETAILS

Organizations demonstrate high PETs awareness but limited adoption—only 42% actively use PETs, and just 5% employ advanced techniques. The barriers are structural rather than intentional: cost (37%), limited in-house expertise (21%), concerns about reduced model performance (35%), and critically, legal uncertainty about whether PETs satisfy regulatory adequacy requirements (30%).

The AI Governance Guidelines emphasize embedding regulatory requirements into system design through technical standards, explicitly citing privacy-preserving approaches as techno-legal solutions. PETs embody this principle: they are technical methods for achieving data protection objectives while retaining data utility for AI development. However, absent regulatory recognition or clarity, organizations hesitate to invest in PET infrastructure whose compliance value remains uncertain.

MeitY may issue guidance explicitly recognizing that properly implemented PETs—including differential privacy, federated learning, secure multi-party computation, trusted execution environments, and advanced anonymization techniques—constitute appropriate technical and organizational measures towards achieving reasonable security safeguards under Section 8(5) (reasonable security safeguards) of the DPDP Act.

This recognition could clarify two distinct pathways: First, certain PET implementations, depending on the specific technique and use case, may achieve effective anonymization as defined in Recommendation B, allowing data to exit the scope of DPDP regulation entirely. Second, even where PETs do not achieve full anonymization, they can constitute reasonable security safeguards under Section 8(5), strengthening data protection while data remains within the DPDP framework.

However, PETs are not a universal solution suitable for all contexts. For resource-constrained organizations, smaller-scale applications, or use cases where simpler measures provide adequate protection, mandating advanced PETs would impose disproportionate costs without commensurate benefit. The Principal Scientific Advisor's white paper on Strengthening AI Governance through Techno-Legal Framework (2026) recognizes this principle. It notes that "larger deployments touching more citizens and or deployments associated with higher perceived risks should have advanced levels of governance, transparency, and technical controls," while acknowledging that the scale of deployment and the level of perceived risk should determine the degree of technical controls required. (Principal Scientific Advisor, 2026). The guidance could therefore be outcome-focused: organizations must demonstrate through testing and documentation that their chosen approach—whether advanced PETs, basic privacy measures, or a combination—achieves adequate privacy protection for their specific context, risk profile, and intended use. This recognizes that as articulated in Recommendation B, full anonymization is neither technically feasible nor strictly necessary in all cases, and that proportionate privacy protection can be achieved through various means appropriate to different organizational capabilities and use case requirements.

Section 17(2)(b) of the DPDP Act exempts processing of personal data for "research, archiving or statistical purposes" if carried out in accordance with prescribed standards. MeitY could clarify

that this exemption applies to developing and testing Privacy-Enhancing Technologies, enabling organizations to use personal data to benchmark different PET approaches under appropriate governance safeguards.

Section 17(2)(b) of the DPDP Act exempts processing of personal data for "research, archiving or statistical purposes" if carried out in accordance with prescribed standards. MeitY could clarify that this exemption applies to developing and testing Privacy-Enhancing Technologies, enabling organizations to use personal data to benchmark different PET approaches under appropriate governance safeguards.

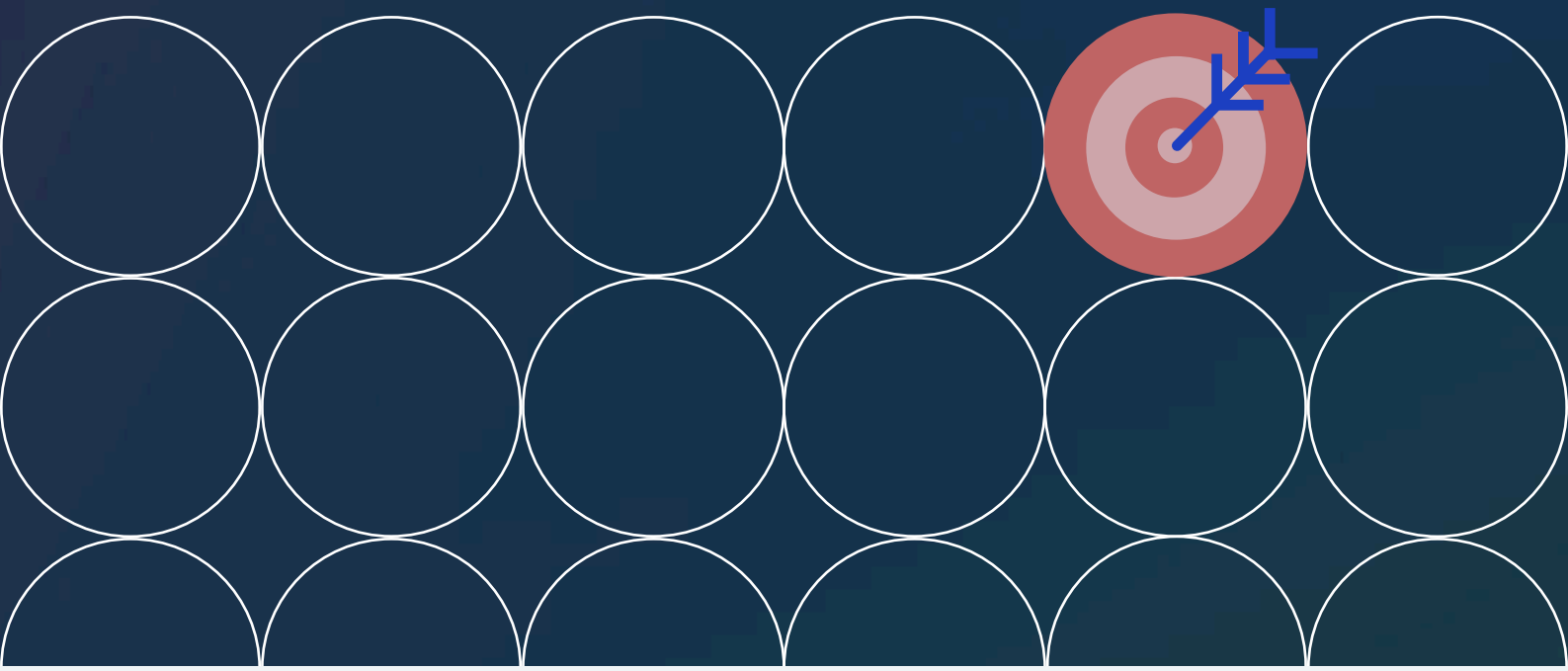
This addresses a practical challenge: organizations need to test whether specific PET techniques effectively protect personal data in their use cases. However, testing PETs requires processing personal data, yet organizations hesitate to do so without certainty that their implementation will be deemed adequate. By clarifying that PET research and development qualifies as legitimate research under Section 17(2)(b), this interpretation creates a compliant pathway for responsible experimentation with personal data while developing privacy-protective capabilities.

For organizations ready to deploy PET-enabled AI systems in production, regulatory pathways could enable pilots under supervisory oversight. The AI Governance Guidelines explicitly recommend regulatory sandboxes; MeitY, working with sectoral regulators, could establish sandbox programs where organizations can test AI applications using novel PET approaches. These sandboxes serve a distinct purpose from the research exemption: they focus on validating deployed systems rather than developing techniques, and they collect empirical evidence on PET effectiveness in real-world Indian contexts that can inform subsequent technical standards.

Technical standards and implementation guidance for PET deployment could be developed, working with standards bodies and research institutions to operationalize different PET techniques for various AI use cases. This includes specifying appropriate privacy parameters for different risk contexts (such as differential privacy epsilon values), methods for documenting how PETs affect model accuracy and fairness, and testing approaches that validate PET effectiveness. For smaller organizations lacking dedicated privacy engineering teams, implementation guidance could provide actionable roadmaps, ranging from prescriptive checklists for common scenarios to principles-based frameworks for more complex deployments.

Addressing cost and capacity barriers will require coordinated support from both government and industry to catalyze PET adoption, particularly for MSMEs, startups, and organizations with limited resources. MeitY has supported deep-tech innovation through initiatives like the SAMRIDH scheme, and the IndiaAI Mission has been backing AI startups, demonstrating pathways for targeted support that can accelerate ecosystem development.

Conclusion



The Open Loop India Program 2025 examined how organizations developing AI systems navigate data protection requirements under the DPDP Act, with a focus on anonymization practices, privacy-enhancing technologies, and compliance pathways for AI-specific use cases. India's AI ecosystem demonstrates strong readiness to align with clear regulatory frameworks. Organizations across sectors showed proactive adoption of privacy safeguards and expressed a clear commitment to responsible data practices. The research revealed not resistance to compliance, but a need for AI-specific clarity on how existing data protection principles apply to iterative model development workflows.

India has taken important steps to address these needs. The Digital Personal Data Protection Rules notified in 2025 operationalize the DPDP Act's framework, while the AI Governance Guidelines released in late 2025 propose institutional mechanisms for coordinated AI governance. MeitY's efforts through initiatives such as the IndiaAI Mission, as well as the establishment of bodies like TPEC and AISI, demonstrate a commitment to building robust AI infrastructure alongside regulatory clarity.

> The three recommendations in this report build on this foundation. Clear anonymization guidance, explicit recognition of the appropriate adoption of privacy-enhancing technologies as compliance tools in certain circumstances, clarity on the publicly available data exemption, and an interim exemption for AI model training together provide the regulatory certainty organizations need to innovate responsibly.

Beyond these targeted interventions, organizations also highlighted the need for high-quality domain-specific datasets and inclusive compliance approaches that account for diverse populations. Addressing these areas will require sustained investment in data infrastructure and continued stakeholder engagement. Implementing the recommendations will require collaboration across government bodies, industry, civil society, and academia to develop technical standards that balance innovation with strong data protection.

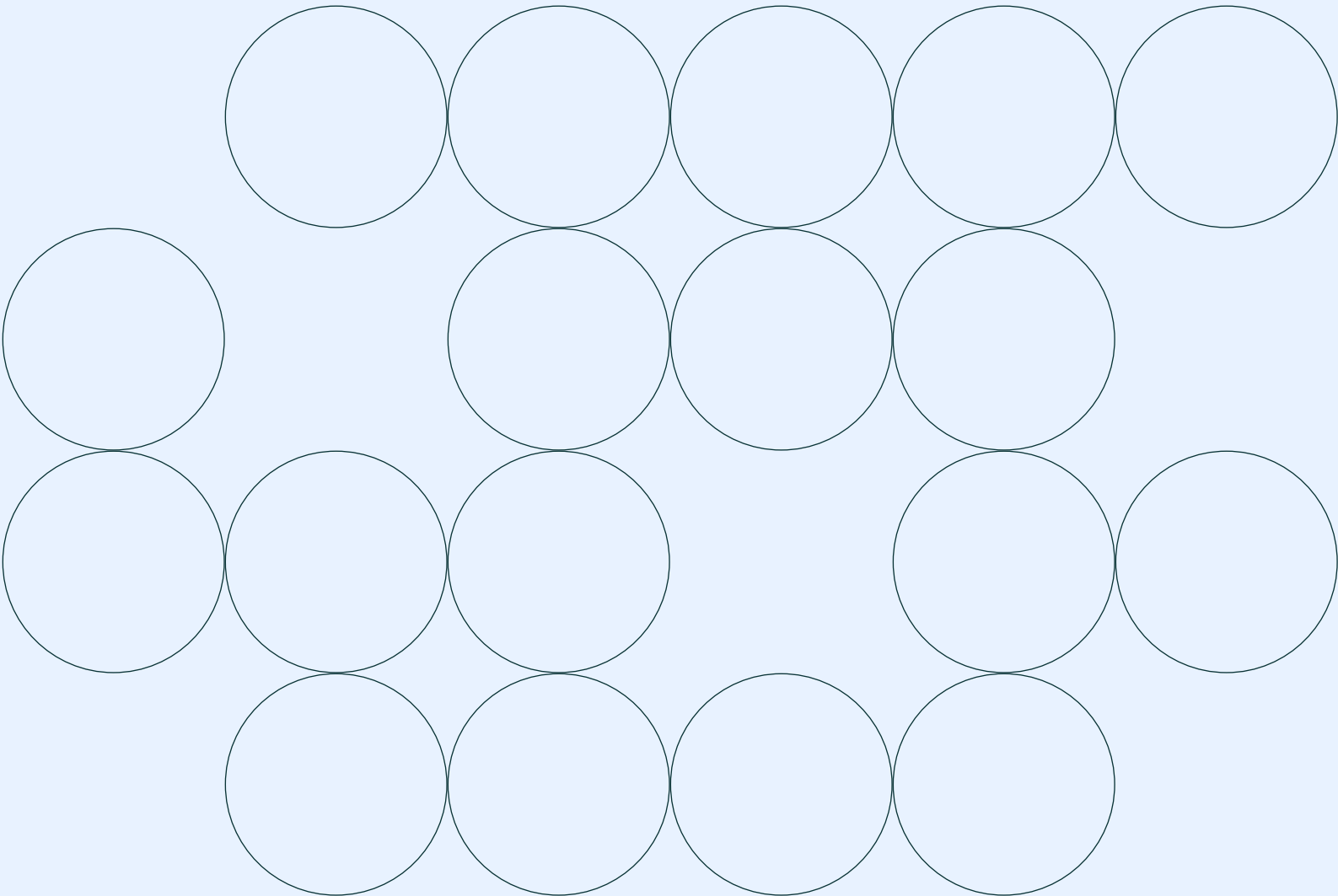
The path forward depends on sustained engagement between regulators and the AI ecosystem. By working together, India can create an environment in which AI technologies flourish while maintaining robust privacy protections. As one of the world's largest and most diverse digital economies, India's approach to balancing data protection with AI innovation offers valuable lessons for other emerging economies facing similar challenges. This collaborative approach will be essential as India continues to develop its AI governance architecture and positions itself as a leader in responsible AI development.

References

1. Advisory Guidelines on the PDPA for Selected Topics (2022). Available at: https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/advisory-guidelines/ag-on-selected-topics/advisory-guidelines-on-pdpa-for-selected-topics-310322.ashx (Accessed: January 2, 2026).
2. AEPD Guidance on Anonymization (2021). Available at: https://www.edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymization_en_5.pdf (Accessed: January 2, 2026).
3. Bagdasaryan, E. and Shmatikov, V. (2019) "Differential Privacy Has Disparate Impact on Model Accuracy." arXiv. Available at: https://doi.org/10.48550/arXiv.1905.12101.
4. Bansal, A. (2025) "MeitY Backs DPIIT's Mandatory AI Copyright Licensing Framework," MEDIANAMA, 19 December. Available at: https://www.medianama.com/2025/12/223-meity-mandatory-ai-copyright-licensing-tech-industry-lawful-access/ (Accessed: January 12, 2026).
5. Bukaty, P. (2021) The California Privacy Rights Act (CPRA) – An implementation and compliance guide. Available at: https://www.jstor.org/stable/j.ctv1kv1d14 (Accessed: January 2, 2026).
6. CARICOM Community (CARICOM) Secretariat Data Protection and Privacy Rules (2020). Available at: https://caricom.org/wp-content/uploads/CCS-Data-Protection-and-Privacy-Rules-GC.pdf (Accessed: December 22, 2025).
7. Court of Justice of the European Union (2025) EDPS v SRB. Available at: https://curia.europa.eu/juris/document/document.jsf;jsessionid=8A02C6116F25EADE8061D10C890960FD?text=&docid=303863&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=1373624 (Accessed: January 2, 2026).
8. Office of the Principal Scientific Adviser to the Government of India (2025) Democratising Access to AI Infrastructure. Available at: https://www.psa.gov.in/CMS/web/sites/default/files/publication/WP_Democratising%20Access_V3.0_29122025A.pdf (Accessed: January 12, 2026).
9. Data Security Council of India (2023) Anonymization Framework - The Future of Data Protection in India. Available at: https://www.dsci.in/files/content/knowledge-centre/2023/Part%20II-Anonymization%20Framework-The%20Future%20of%20Data%20Protection%20in%20India%20Report%202023.pdf (Accessed: December 22, 2025).
10. European Commission (2025) Digital Omnibus Regulation Proposal | Shaping Europe's digital future. Available at: https://digital-strategy.ec.europa.eu/en/library/digital-omnibus-regulation-proposal (Accessed: December 22, 2025).
11. European Union (2016) General Data Protection Regulation (GDPR) – Legal Text, General Data Protection Regulation (GDPR). Available at: https://gdpr-info.eu/ (Accessed: January 2, 2026).
12. European Data Protection Board (2025) Guidelines 01/2025 on Pseudonymisation. Available at: https://www.edpb.europa.eu/system/files/2025-01/edpb_guidelines_202501_pseudonymisation_en.pdf (Accessed: December 22, 2025).
13. ISO/IEC 20889:2018(en), Privacy enhancing data de-identification terminology and classification of techniques (no date). Available at: https://www.iso.org/obp/ui/en/#iso:std:iso-iec:20889:ed-1:v1:en (Accessed: January 12, 2026).
14. Mohamed, B. (2024) "Five ways in which the DPDPA could shape the development of AI in India - Future of Privacy Forum," https://fpf.org/. Available at: https://fpf.org/blog/five-ways-in-which-the-dpdpa-could-shape-the-development-of-ai-in-india/ (Accessed: January 12, 2026)
15. Naveed, H. et al. (2024) "A Comprehensive Overview of Large Language Models." arXiv. Available at: https://doi.org/10.48550/arXiv.2307.06435.
16. Personal Data Protection Act 2012 - Singapore Statutes Online (no date). Available at: https://sso.agc.gov.sg/Act/PDPA2012 (Accessed: January 30, 2026).
17. Utah State Legislature (2022) Utah Consumer Privacy Act, Utah Code Ann. § 13-61-101 et seq. (effective 31 December 2023)
18. Principal Scientific Advisor (2025) Report on AI governance guidelines development, IndiaAI. Available at: https://indiaai.gov.in/article/report-on-ai-governance-guidelines-development (Accessed: December 22, 2025).
19. Sinha, A. (2024) "The Many Questions About India's New AI Advisory," Tech Policy Press, 6 March. Available at: https://www.techpolicy.press/the-many-questions-about-indias-new-ai-advisory/ (Accessed: January 2, 2026).
20. Principal Scientific Advisor (2026) Strengthening AI Governance through Techno-Legal Framework, Available at: https://psa.gov.in/CMS/web/sites/default/files/publication/AI-WP_TechnoLegal.pdf (Accessed: January 30, 2026).

AI Innovation, Effective Anonymization & the DPDP Act

Annexes



Annex I: Glossary (key concepts)

This report draws on various technical concepts relevant to GenAI applications. The following table provides definitions and explanations of these concepts:

Sr No.	Fundamental concept	Explanation
1	Anonymization	Anonymization is the process of removing or altering information that can be used to identify a specific individual in a manner that is very difficult to reverse, and where the stated goal is to make reversal impossible (DSCI, 2023). Anonymization reduces re-identification risk, especially when sharing beyond original context. Anonymization techniques fall on a spectrum where irreversibility and usability share an inverse relationship – the more highly anonymized and irreversible the data is, the less useful it will be. Conventional methods often anonymize to the point of weakening correlations in datasets, rendering data unusable for machine learning models, particularly in the case of GenAI. For instance, applying Differential Privacy (DP) adversely impacts model accuracy, especially in the case of underrepresented groups – showing more accuracy with white faces than with black faces, for instance. (Bagdasaryan and Shmatikov, 2019). Anonymization must be understood as risk minimization rather than risk elimination. Data protection frameworks should define anonymization based on reasonable risk thresholds (e.g., 'reasonably likely means') rather than impossible standards of absolute irreversibility, particularly where such standards would render data unusable for legitimate purposes.
2	Identifiability	Identifiability refers to the ability to distinguish an individual from others (DSCI, 2023). An individual is deemed 'identified' or 'identifiable' if their personal information, either alone or in combination with other data, can pinpoint them. The spectrum of identifiability starts from data that is fully identifiable on one end. This includes data that directly points to a specific person, such as a name or Aadhaar number. In the middle, there is de-personalized data, in which direct identifiers have been removed or encrypted, but the data is still about an individual and could theoretically be re-identified. On the other end is anonymized data, where identifying attributes have been irreversibly removed or transformed.
3	Privacy Enhancing Technologies (PETs)	For the scope of this report, we define Privacy-Enhancing Technologies (PETs) as technical and organizational tools, processes or methods that enable personal or sensitive data to be collected, processed, analyzed or shared while preserving individual privacy. This includes methods for reducing identifiability, limiting access, applying anonymization, encryption or secure computation.

Sr No.	Fundamental concept	Explanation
4	Pseudonymization	Pseudonymization involves the replacement of direct identifiers with pseudonyms, tokens, or codes, while allowing re-identification via separately a stored key or linkage table. Regulatory regimes like the GDPR have classified such data as personal, given the feasibility of reidentification (European Data Protection Board, 2025). However, recent jurisprudence has upheld a more contextual understanding of personal data, clarifying that pseudonymized data does not have to consist of personal data in all circumstances (Court of Justice of the European Union, 2025). In some situations, pseudonymization may effectively prevent persons other than the data controller from identifying the data subject. Whether data is identifiable needs to be assessed (a) at the time of collection, and (b) from the perspective of the controller.
5	Training Data for GenAI	GenAI models require specific datasets and retraining to evolve into more sophisticated, domain-specific, expert models that can be applied for specific use cases. In industries such as healthcare, finance, and insurance this means exposure to personal data that can teach them about specific use cases. The ability to generate relevant and fluent content comes from exposure to high-quality, diverse data. Several GenAI models are pre-trained on large-scale datasets scraped from the open internet through web crawling and user interactions, and many are also trained using proprietary datasets. They include a mix of structured and unstructured information, and contain personal or sensitive information from public social media profiles, blogs or forums. Thus, some of the data in pre-training datasets can be publicly available personal data, and within it, a smaller subset could be sensitive personally identifiable information (PII). However, given the prevailing ambiguity on what constitutes “publicly available personal data” in the Indian context, it remains unclear if such data scraped from the web would be exempt under Indian law. LLM training pipelines sometimes incorporate heuristic-based filters to identify and remove or anonymize data, such as names, addresses, and phone numbers, mitigating the risk of models memorizing or reproducing private details (Naveed et al., 2024). However, in the absence of regulatory clarity, these remain inconsistent practices.

Annex II: Research methodology

This report presents the findings and recommendations generated by the Open Loop India Program 2025, which used mixed-methods research and analyses. The Program sought to understand how organizations in India navigate issues such as data privacy, anonymization, and compliance under the DPDP Act, while using personal data to develop or deploy AI systems. The report identifies challenges and opportunities faced by these companies and offers insights to assist policymakers in developing additional guidelines that can support GenAI use and development. .

The program centers on three core questions:

> **How do organizations collect, manage, and safeguard personal (and non-personal) data in AI workflows?**

> **How and when do they utilize PETs (including anonymization)?**
>
> - What operational challenges and compliance risks exist under the current regulatory framework?

> **What type of regulations and related measures can best enable responsible and effective AI innovation across India?**
>
> - Which facets of the DPDP Act and Rules require further adaptation to AI, which can spur greater innovation and mitigate compliance risks?

Participant recruitment & cohort design

To ensure sectoral diversity and representation across the AI ecosystem, the program reached out to organizations from different sectors, and of different sizes and positions along the AI value chain. The cohort comprises 44 organizations across 13 sectors.

Survey design and administration

A structured online survey was designed to capture both quantitative trends and qualitative insights. The survey comprised four sections:

1. **Profiles**
   The nature and size of organizations and their data governance and use practices, including data sources, handling, and protection mechanisms.
2. **Compliance**
   Compliance readiness and experience Familiarity with and readiness for compliance under the DPDP framework.
3. **Data Protection Practices**
   The knowledge and use of PETs and anonymization practices.
4. **Challenges**
   Ongoing challenges faced and additional regulatory guidance required.

The survey comprised multiple-choice, ranking, and open-ended format questions that captured measurable trends and specific organizational practices. Participants could collaborate internally, for instance, between policy, legal, managerial, and technical teams, to complete the survey.

Survey responses were analyzed using quantitative and qualitative methods. In most questions, respondents were allowed to select multiple options to reflect hybrid business models and layered data practices; therefore, the total count (n) for individual questions may exceed the number of companies, unless otherwise stated.

Follow-up interviews

To complement the survey data, the program included follow-up, semi-structured interviews with 14 participants. These interviews explored their practices in greater depth, including anonymization methods, compliance, and regulatory or technical suggestions for the sector, and elicited their perspectives and insights drawn from their operational experiences.

Secondary research & expert consultations

The program also conducted extensive secondary research and analysis of relevant Indian and global policy frameworks, which were accompanied by consultations with technical experts and policy and legal practitioners to test emerging insights and validate findings.

Limitations of the research

While this report offers valuable insights into how Indian organizations approach data privacy, anonymization, and compliance in AI systems, its findings should be interpreted within certain methodological limitations:

1. **Sample size and representativeness:** The survey reflects responses from **44 organizations**, which provides directional insights but limits the report's ability to generalize. The purposive sampling strategy was aimed at ensuring diversity across sectors and roles in the AI value chain, however the results may not represent the full spectrum of India's AI ecosystem.
2. **Self-reported data:** All findings are based on self-reported surveys and interview responses. These reflect participants' stated practices and perceptions rather than independently verified data.
3. **Cross-sectional design:** The study captures organizational practices and perceptions at a single point in time (mid-2025). As AI governance and compliance landscapes evolve, some findings may change.
4. **No size-based disaggregation:** The report does not differentiate responses by organization size (e.g., startups, small and medium enterprises, or large enterprises). This may obscure variations in compliance maturity or resource constraints that could influence responses.
5. **Potential social desirability bias:** Considering the regulatory sensitivity of the topic, some responses may be influenced by reputational or compliance-related considerations.

Annex III: Policy Background

Annex 3(1): Notable provisions of DPDP Act and rules framework

The DPDP Act stands on two pillars of **notice** and **consent**, which require companies to provide notice and obtain consent before collecting, using, or processing personal data. It places protections, such as using reasonable security safeguards (Section 8(5)), erasing data once used (Section 8(7)), and addressing grievances (Section 8(10)).

Participating organizations from the Open Loop India Program 2025 have indicated gaps in data regulation for GenAI innovation. The following is a detailed breakdown of the specific legal provisions pose regulatory uncertainty for GenAI innovation:

(I) The potential requirement of explicit consent across iterative stages of data processing for GenAI training can create substantial operational difficulties and costs

Section 4 of the DPDP Act lists two grounds for processing personal data - through consent, or for certain legitimate uses. The DPDP Act does not explicitly state whether AI model training constitutes processing under consent (Section 6) or could qualify under certain legitimate uses (Section 7). Organizations are unsure of which legal basis applies to different AI workflows, particularly for iterative data reuse across training, fine-tuning, and improvement stages.

Section 10 of the DPDP Act enables the notification of certain data fiduciaries as significant data fiduciaries, based on the volume and sensitivity of the data they process, among other factors. This creates additional obligations, such as periodic audits and data protection impact assessments. Without a specific policy guidance on whether processing data for GenAI would attract additional obligations under Section 10 for data fiduciaries, innovators are at risk of being underprepared for meeting audit and associated requirements. The newly notified DPDP Rules restrict significant data fiduciaries from cross-border data flows for some personal data (Rule 13(4)), making this a foreseeable risk for several GenAI business operations.

(II) Permissibility of alternative pathways to use data for training purposes is uncertain under the DPDP Act

The DPDP Act does not define anonymization or provide a certain "effective anonymization" standard that organizations can confidently achieve through de-identification techniques. This keeps the status of data in a regulatory grey area, restricting AI-innovation, even with privacy protections.

Section 17(2)(b) of the DPDP Act exempts processing of personal data from regulation if such processing is "necessary for **research, archiving or statistical purposes** if the personal data is not to be used to take any decision specific to a Data Principal and such processing is carried on in accordance with such standards as may be prescribed." Companies are not sure whether the interpretation of "research" extends to research required for PET innovation and commercial model training, or if it is limited to academic research in the narrowest sense of the term. **This is reflected in their responses – developing and testing PETs is a risk when the DPDP Act does not exempt personal data for security benchmarking exercises.**

The Explanatory Note to the DPDP Act suggests that this provision facilitates "necessary data processing for academic and policy research." However, it neither strictly limits processing to academic and policy research in the narrowest sense, nor does it carry binding obligation.

(III) Current Gap: Lack of Incentives for PET Adoption in AI Deployment	The DPDP Rules mention data security measures, such as encryption, obfuscation/masking or virtual tokens mapped to personal data, **which hint at anonymization and pseudonymization techniques (Rule 6)**. But they do not clarify: (a) whether the expectation is to de-identify personal data as a security measure, (b) whether there are **specific contexts that require more intensive de-identification or even anonymization,** or (c) whether data fiduciaries will continue to be liable under the DPDP Act **even after the personal data has lost its identifiable characteristics. This creates a "Schrödinger's data" problem: data that has been anonymized exists in regulatory limbo – organizations cannot be certain if DPDP Act obligations cease to apply, creating hurdles for downstream uses.**

Annex 3(2): India's AI policy landscape and present-day considerations

India does not have a specific legislation on AI and its AI policy outlook can be gauged from its policy instruments and literature on AI adoption, some of which are sectorally in focus. Notably, India's National Strategy for Artificial Intelligence (2018) recognizes the regulatory ambiguity on anonymization as a barrier to AI-led innovation. **Subsequent attempts at regulating AI have addressed sectoral fragmentation and foundational principles, but haven't translated into enforceable standards yet:**

A MeitY Advisory on AI deployment in the run up to the elections in March 2024 stipulated that all under-tested or unreliable AI models must get approval from the government before being deployed. A subsequent clarification issued through an official social media post on X (formerly Twitter) by the Minister of State mentioned that the Advisory would apply only to "significant platforms" (Sinha, 2024). While they are helpful clarifications, formal codified guidance would provide stronger clarity and enforceability.

An Advisory Group chaired by the **Principal Scientific Advisor (PSA)** was constituted to develop an "AI for India - Specific Regulatory Framework." Under its guidance, a sub-committee released a Report on AI Governance Guidelines Development in January 2025 for public consultation (Office of the Principal Scientific Advisor to the Government of India, 2025). It identified a gap in the application of the cybersecurity framework (CERT-IN Rules, Cybersecurity Directions 2021, DPDP Act, NCIIPC Rules) in the context of AI systems. The Report proposed regulating AI and recommended: establishing a **technical advisory body** comprising multidisciplinary experts from academia and industry; building an AI incident database for evidence-based harms mitigation; voluntary transparency commitments from the industry across the AI ecosystem; using technological measures to address AI-related risks; and forming a MeitY subgroup to suggest specific measures under the proposed legislation Digital India Act, to harmonize the legal framework and grievance redressal. Crucially, while the report emphasizes institutional and procedural safeguards, regulatory clarity on foundational issues (such as, standards for anonymization and data use for AI model training) are essential for enabling responsible innovation at scale.

In August 2025, the **Reserve Bank of India (RBI) released its FREE-AI (Framework for Responsible and Ethical Enablement of Artificial Intelligence) Committee Report.**

- **Recommendations for AI Adoption:** The report lays down seven foundational principles and 26 targeted recommendations across innovation enablement and risk mitigation. The Report has a **positive approach towards AI adoption**, specifying how AI can achieve financial inclusion through its diverse use cases. For instance, AI can assess creditworthiness from non-traditional data sources, thus onboarding thin files or new-to-credit borrowers. DPIs such as Know Your Customer (KYC) and Unified Payments Interface (UPI) can leverage AI to improve functioning.
- Identification of Risks and Gaps: However, the report also highlights risks associated with AI adoption, particularly increased financial vulnerabilities, ethical, privacy and cybersecurity concerns, third party risks, operational and model risks. **The report notably highlights institutional risks – that most entities lack policies for data management for training AI, and that the industry has expressed a need for regulations on AI adoption, guidelines on data privacy, algorithmic transparency, the use of external Large Language Models (LLMs), cross-border data flows etc.**

The **India AI Governance Guidelines** released in November 2025 present a flexible, **pro-innovation approach** to governing AI. They propose a **graded liability system** for AI, depending on the specific function performed, the level of risk posed, and the extent to which due diligence was observed. Specifically, they recommend targeted legislative amendments to encourage innovation, and **regulatory sandboxes with reasonable legal immunities** to encourage building cutting-edge AI technologies. The guidelines specifically acknowledge the regulatory uncertainties surrounding data protection and AI governance, and highlight potential pathways towards techno-legal design to redress some confusions:

- **Applicability of DPDP Act for AI:** The guidelines present an assessment of the DPDP Act that might offer some direction - that personal data usage without obtaining user consent for training AI models is governed by the DPDP Act. At the same time, the **guidelines acknowledge ambiguities in the law** - its applicability of exemptions on publicly available personal data, compatibility of collection and purpose limitation principles with modern AI systems, role of consent managers, notice and legitimate exceptions in the context of multi-modal AI.
- **Guidelines endorse techno-legal approach:** The guidelines propose inbuilt "compliance by design" features to scale innovation while mitigating risks. The guidelines emphasize the importance of prior testing of techno-legal measures, and list privacy preserving tools for AI development as a techno-legal measure. The guidelines recommend modifying the Data Empowerment and Protection Architecture (DEPA) for AI training.
- **AI Governance Group and Voluntary frameworks:** The guidelines recommend establishing an AI Governance Group to coordinate AI policy development and align AI governance frameworks with national priorities, besides voluntary frameworks such as industry codes of practice, technical standards and self certifications, aligned with sectoral assessment of risks.

Annex 3(3): Common internationalstandards for anonymization policy

Participant-companies mentioned some globally accepted anonymization standards to indicate the level of regulatory clarity they seek in determining the anonymization spectrum. Below is a snapshot that captures the diversity in anonymization spectrum:

Across jurisdictions around the world, a risk-based approach to anonymization has been the most common approach, and has operated with varying thresholds. For instance, The **General Data Protection Regulation (GDPR) mandates considering "all the means reasonably likely to be used"** for identification, factoring in time, cost, and technology, and considers effective anonymization to be fulfilled when an individual cannot be singled out, linked to a record, or have information inferred about them (European Union, 2016). In addition, it is important to note that the GDPR has created many barriers for companies in the EU, as a result of which the European Commission is now undertaking a process of revision, consolidation and simplification of the entire Digital Rulebook, and should not be considered the "gold standard". Similarly, the **California Privacy Rights Act (CPRA)** requires that information **"cannot reasonably be used" to identify or link to a person**, and promotes procedural safeguards such as public commitments and contractual obligations to commit to anonymization (Bukaty, 2021). **Singapore's Personal Data Protection Commission (PDPC) sets a "serious possibility of re-identification threshold"**, and any possibility of data identification by a motivated person deems the data non-anonymous (Singapore PDPC, 2022).

Under **Convention 108+** an individual is not considered "identifiable" if his or her identification would require **unreasonable time, effort or resources**. What constitutes "unreasonable time, efforts or resources" should be assessed on a case-by-case basis, such as depending on the cost, the benefits of such an identification, the type of controller, the technology used, etc. Further, **technological and other developments may change what constitutes "unreasonable".**

The **HIPCAR** project (Enhancing Competitiveness in the Caribbean through the Harmonization of ICT Policies, Legislation and Regulatory Procedures) does not specifically mention anonymized data. However, its data protection and privacy rules stipulate **anonymization as a method of disposal or erasure of personal data** which no longer serves the original purpose of collection, effectively exempting data from regulation (HIPCAR PDPR, 2020). Another threshold is presented by the **Spanish Data Protection Agency** (Agencia Española de Protección de Datos or " AEPD "), which specifically clarifies **context as a determinative factor** for anonymization in its policy guidance, while broadly aligning with the GDPR on anonymization. An example shared by the AEPD on context-dependence highlights how re-identification risks can vary across thresholds depending on the availability of datasets in different circumstances. For instance, details of taxpayers' personal data are publicly available in Sweden, but not in Spain. Hence, even if the same procedure is used to anonymize datasets of Swedish and Spanish citizens, the re-identification risks could differ. The Spanish Data Protection Agency acknowledges that **sometimes re-identification risks cannot be minimized beyond a threshold where the data is still usable** (AEPD, 2021).

In each case, a clear standard for anonymization – in the form of "reasonably likely means" or "cannot reasonably be used" standard – has been vital to regulate data processing in a manner that enables AI innovation. By contrast, India does not specify a threshold for anonymization either form, leaving AI developers and users in the dark and unable to responsibly leverage datasets to innovate.

Beyond these varying thresholds of risk-based anonymization, Europe is seeing a broader regulatory shift in its data protection framework, with the **European Digital Omnibus** proposing several amendments to the digital legislative framework (European Commission, 2025). Some of the concerns that the Digital Omnibus raises about the GDPR, especially pertaining to smaller organisations that perform non-intensive, low risk data processing operations but face heavy obligations under the law, **are echoed by some members of the Open Loop India Cohort through intensive surveys and interviews**. The Digital Omnibus lends necessary clarity to the notion of personal data, distinguishing it from pseudonymized data in the light of the recent CJEU verdict. It also lends definitional clarity to scientific research, extending its scope to systematic research and development to industry settings, and adding that (a) further processing for scientific purposes is compatible with the initial purpose of processing data and (b) scientific research is a legitimate interest. Similar clarity from the regulatory framework in India on valid thresholds for processing data for AI training would equip the industry, especially small and medium sized enterprises, to innovate responsibly.


Open Loop
Meta