# Data Empowerment and Protection Architecture (DEPA) for AI Training:

## *A Discussion Paper*




**Naavi**

(Vijayashankar Nagaraja Rao)

Revised edition | 26 November 2026

*In the light of AIGSI Version 3.0*

## Data Empowerment and Protection Architecture (DEPA) for AI Training: A Discussion Paper

Naavi (Vijayashankar Nagaraja Rao)

Revised edition, 26 November 2026, in the light of AIGSI Version 3.0

(First edition presented at IDPS 2026, Bengaluru, 21 November 2026)

> **About this revised edition.** The first edition of this paper, presented at IDPS 2026 in Bengaluru, set out how AIGSI Version 2.0 could fill the governance gaps in DEPA for AI Training. Since then FDPPI has prepared AIGSI Version 3.0 as a discussion paper for its proposed Special Interest Group (SIG) on AI Governance. Version 3.0 deals directly with several issues this paper raises: where the liability chain ends, what happens when roles blur, who becomes a Model Owner when a model is fine-tuned, fairness and collective harm, copyright, release controls, and a single audit that produces a CERT-In opinion, a DPDPA opinion and an IT Act opinion.
>
> This edition rewrites Part E around AIGSI Version 3.0, adds an AIGSI column to the table of open issues in Part C, adds cross-references in Part D, adds recommendations to Part F, and adds Questions 36 to 41 to the Annex. **Questions 1 to 35 keep their original numbers**, so responses already sent quoting those numbers remain valid. The analysis of the DEPA handbook in Parts A to D is otherwise unchanged.

### Preface

On 31 October 2026, MeitY released the *Handbook on Data Empowerment and Protection Architecture (DEPA) for AI Training* at a roundtable in Bengaluru. Several dignitaries spoke at the launch.

**Shri S. Krishnan, Secretary, MeitY:**

> "India's AI ambitions must be matched by the ability to use data responsibly and securely. The DEPA Training Framework enables trusted data collaboration at scale, unlocking access to large-scale datasets for AI training while safeguarding privacy, commercial confidentiality and institutional control. MeitY is fully committed to taking this framework forward to advance AI innovation in the country."

**Prof. Ajay Kumar Sood, Principal Scientific Adviser to the Government of India:**

> "Under DEPA, the model travels to the data, rather than the data travelling to the model. This enables trusted, multi-party and cross-border data collaboration, accelerating scientific discovery and helping solve problems that no single institution can address alone. DEPA is India's contribution to global science diplomacy, advancing the efforts of institutions such as the World Bank and the United Nations to shape a more inclusive and responsible global AI ecosystem."

**Shri Sharad Sharma, Co-founder, iSPIRT Foundation:**

> "Today marks the whole data economy (of many pillars) coming into place. This is easily a more significant moment since February 2016, when the Unified Payment Interface (UPI) was announced. India has the most distinctive, safe and flexible data economy in the world. Because of this, India now has a place in the AI stack like never before."

**Shri Anjani Rathor, Group Head – Digital Banking and Customer Experience, HDFC Bank:**

> "Trusted data collaboration can unlock significant value for AI in banking. Our experience with DEPA Inference has demonstrated the promise of secure data collaboration. We see a compelling opportunity to build on this experience through DEPA Training."

**Shri Kris Gopalakrishnan, Co-founder, Infosys and Chairman, Axilor Ventures:**

> "India has an opportunity to lead in safe AI by enabling trusted access to data for research and innovation. DEPA Training can help researchers harness diverse datasets to develop more robust AI models, improve collaboration across multiple teams, and advance scientific discovery, while safeguarding the rights and interests of data custodians and contributors."

## Why this paper

The 41-page handbook acknowledges that the initiative was pioneered by the iSPIRT Foundation, a body of industry leaders who earlier ran product initiatives within NASSCOM. iSPIRT has now proposed a new DEPA architecture that lets AI model trainers access bulk data from Indian companies and the Government. The framework is referred to as DEPA-AIT (DEPA for AI Training).

In effect, the framework opens access for AI model developers to valuable personal data, which is a sovereign Indian asset. Access is not reserved for Indian developers. So this is not an initiative to build indigenous AI capability. It may instead end up enriching OpenAI, Anthropic, Google, Meta and DeepSeek with data from the Indian Government and Indian companies.

The framework will create a facility where data from different sources in India is collected and made available for AI model training. Prudent Indian companies may hesitate to share their data on such a platform, but the Government may well do so. DEPA-AIT may therefore become mainly a channel for moving Government-held personal data of Indian citizens to the leading AI model developers in the US.

The Secretary, MeitY has invited "actionable recommendations" for refining and co-creating this framework, which is presented as a "Digital Public Infrastructure" for the benefit of all. This paper responds to that invitation.

Naavi presented the first edition of this discussion paper at IDPS 2026, Bengaluru on 21 November 2026, and presents this revised edition at IDPS 2026, Hyderabad on 28 November 2026, to invite comments from fellow professionals.

Comments may be sent to naavi@fdppi.in with the subject line "DEPA for AI Training". Participants are requested to respond to the questions in the Annex: Questions for Discussion at the end of this paper, quoting the question number.

**Naavi**

## Part A: The core concerns

### The DPDPA context

The backdrop to this document is the DPDPA 2023, whose main obligations come fully into force in about six months. Under the Act, a Data Fiduciary may process personal data only with the specific consent of the Data Principal, unless a recognised exception applies. Consent is purpose-specific. The same data cannot be used for another purpose, including selling or otherwise sharing it for AI model training.

The Act permits a few uses of personal data without consent as "legitimate uses" under Section 7. Separately, Section 17(2)(b) exempts processing for research, archiving or statistical purposes, subject to prescribed conditions. Data that is genuinely anonymised falls outside the Act altogether.

Any other use of personal data, whether direct or through a processor, is prohibited and could lead to penalties that may add up to around ₹850 crore in total. Breaches may also result in imprisonment of company officials for their negligence and in compensation payable to affected Data Principals.

Given this exposure, the Government knows that no private company can easily be persuaded to share its data for AI training. The Government itself may share data, but it too risks liability under DPDPA 2023.

### The "Data Deadlock" and the pressure to break it

The Government has therefore recognised a "Data Deadlock": rich data sits with industry and public bodies but cannot be used for AI training. Industry is now pressing the Government to break this deadlock by creating a legal channel through which the personal data of Indian citizens can reach AI model developers.

In Naavi's view, DEPA-AIT is a framework that facilitates such personal data transport.

DPDPA 2023 is already before the Supreme Court on questions of constitutional validity. At this time, it is surprising that the Government is yielding to commercial pressure to dilute the Act through manoeuvres such as DEPA for AI Training. If this is raised before the Court as evidence of the Government's bad faith in protecting privacy, it could seriously weaken the Government's defence in the cases now pending.

### A "UPI for data", built without visible consultation

The vision is to make this framework a Digital Public Infrastructure like UPI, where organisations contribute data and AI model developers harness it.

The document says it was developed through extensive collaboration between Government, industry and academia. But no details of such consultation are publicly available. Since this is a public project, the record of how it was developed should be obtainable through RTI from the public institutions involved.

### The CIBIL precedent

The possibility that this framework is a backdoor through which AI industry can access and exfiltrate the personal data of 140 crore Indians is glaring. A similar data laundering exercise occurred around 2014, when TransUnion was allowed to take control of CIBIL through a transfer of controlling shares. Control over some 500 million data records worth billions of dollars passed with it, at a throwaway price of around ₹3,800 crore. The current proposal must therefore be viewed with suspicion until proved otherwise.

### Endorsements and the "need to share" narrative

The document carries endorsements from several dignitaries, in an effort to build support for highlighting the "need to share data" for AI innovation. As one foreword puts it, the handbook is "a statement of intent and a positive contribution to the global thinking on data and AI governance."

If the Government's intention were national benefit, it would have recognised that this proposal places a treasure of sovereign data in the hands of international AI companies, from whom Indians will later have to buy services back. It is like the days of the British, when raw materials were shipped from India to England and value-added products imported back.

It is impossible to believe that MeitY, iSPIRT or NASSCOM is unaware of this. The inference is that the Government is yielding to the commercial pressure of the AI lobby led by Meta, Google, Anthropic and OpenAI.

The Comptroller and Auditor General (CAG) recently flagged the need to determine the fair value of data held by PSUs, and instructed auditors to have it recognised during audits. It is therefore surprising that MeitY now proposes to give away a national asset without proper thought.

Had sharing been restricted to Indian companies, there would have been a defensible logic that Indian data enriches Indian AI models. No such restriction is proposed. As drafted, this appears to be a scheme for making Indian data available to foreign AI model developers.

The proposal needs to be understood and debated in this context.

## Part B: What the framework proposes

### Principles and participants

The framework is said to rest on the ADEPTS principles (Accountability, Democratization, Evolvability, Privacy, Transparency and Security), which are meant to establish the foundation of trust for India's data ecosystem.

The two principal participants are:

- **Training Data Providers (TDPs):** institutions such as hospitals, banks and public agencies that hold valuable training data.
- **Training Data Consumers (TDCs):** innovators such as start-ups, researchers and enterprises.

The stated purpose is to bring TDPs and TDCs onto a common techno-legal platform, where secure, private and compliant data exchange and AI modelling can produce data-driven solutions for societal impact.

The framework rests on four core pillars:

1. Data Exchange Platform
2. Electronic Contracts
3. Secure Data Environment (SDE)
4. Differential Privacy

Through this combination of technology, contracts, privacy-enhancing techniques and institutional oversight, DEPA-AIT seeks to overcome the "Data Deadlock".

### The Data Exchange Platform: a data repository

The Data Exchange Platform (DEP) is expected to play the role of a data repository. TDPs provide bulk data, which is anonymised and encrypted and made available for download and use by registered TDCs and AI model builders. Two options are proposed. In the first, the TDP does the anonymisation. In the second, the DEP does it.

It is not clear whether the private sector will adopt the framework and absorb the DPDPA risks. Government departments, however, may be persuaded by MeitY to adopt it. DEPA-AIT then becomes the instrument through which data held by the Government, a sovereign data asset, reaches OpenAI, Anthropic, Meta or Google. It will be interesting to see how the CAG reacts to this proposition.

### The techno-legal approach to governance

DEPA-AIT proposes to bring all critical stakeholders (data providers, AI developers, government representatives, technology specialists and domain experts) into a consortium of market players that would provide oversight for a trustworthy and responsible innovation ecosystem.

It recognises the need for an oversight body to anchor essential functions across the AI development lifecycle:

- setting clear standards for privacy, ethics and security

- registering eligible participants
- overseeing techno-legal mechanisms for the data sharing procedure
- facilitating dispute resolution

### Integration with national data exchanges

The document states that DEPA becomes more powerful when combined with complementary national data exchanges: AI Kosh, API Setu, the NIC API Exchange Platform (NAPIX), the National Data and Analytics Platform (NDAP), the Geospatial Data Infrastructure (GDI) and Open Government Data (OGD). Secure access to these platforms, it says, would enrich AI models across all sectors with vital context.

This indicates an intention to integrate all these data sources into the platform. In Naavi's view, the intention is clear: to put all available personal data of citizens to use for the benefit of global players, so that India ends up eating out of their hands later.

The document does acknowledge the risks under DPDPA, but the acknowledgement lacks credibility.

### Why consent is set aside

The document argues that the earlier DEPA framework and similar architectures, based on consent between the Data Principal and the Data Fiduciary, cannot be followed for AI. It says a different model is needed for sharing personal data of Data Principals with business entities.

Because AI training requires "privacy-preserving pooling" of aggregate data, the current consent-based framework is flagged as insufficient and as the cause of the "Data Deadlock". Having identified the privacy challenges in data sharing, the document goes on to recommend a new architecture.

The proposed Digital Public Infrastructure is meant to move away from bilateral agreements to a collaboration between data providers and AI model developers, supported by enabling participants such as cloud service providers, model providers and certifiers.

### The ecosystem roles

According to the document, DEPA for Training is made up of distinct but interdependent participants, each with a defined role:

1. **Data Principals**, who "may not interact directly with the training network". Their interests are to be protected through an institutional governance structure that operationalises electronic contracts, confidential computing and differential privacy.
2. **Data Collabs**, which bridge the gap between Data Principals and the AI innovation network. These intermediaries, which may be cooperatives, industry associations or research consortia, work with communities to aggregate, standardise and organise data from multiple sources. They "may be responsible for obtaining primary consent where applicable". *Naavi's note: this cuts into the role of the Consent Manager under DPDPA.*
3. **Training Data Providers**, the institutions that source, curate and prepare datasets for AI training. A TDP must apply privacy-preserving transformations and ensure that datasets meet technical and regulatory requirements for de-identification. TDPs include hospitals,

financial institutions and public agencies. They register with a central registry and are bound by its governance standards, enforced through electronic contracts.

4. **Training Data Consumers**, the innovators who access these datasets to develop, refine or evaluate AI models. TDCs must operate within the purpose limitation and ethical conduct defined in their data-sharing agreements, either with Data Collabs that federate datasets for a sector or with individual TDPs. *Naavi's note: these correspond to the research organisations contemplated under DPDPA.*

### Privacy and security by design

The document treats two pillars as the basis of privacy and security by design: rigorous application of **Differential Privacy** and comprehensive enforcement of a **Zero-Trust Security Framework**.

It argues that traditional anonymisation is not sufficient for training AI models, while overly anonymised data loses its usefulness for training. Differential Privacy is therefore recommended. To complement it, a Zero-Trust framework would strictly verify every data access request and protect data at all stages through encryption, with continuous auditing and real-time monitoring.

### Economics

The document claims that the data deadlock costs trillions of dollars in unrealised revenue and productivity. No specific evidence is offered for this figure.

### The proposed workflow

Ministries and other participants are expected to publish a Data Sharing Procedure, and a phased implementation roadmap is proposed. The framework describes a five-stage process:

1. **Data registration and encryption.** The TDP registers and catalogues its dataset, applies anonymisation and encryption, and attaches metadata on origin, sensitivity and access restrictions.
2. **Data request and contract validation.** The TDC submits a request for a specific use case. An electronic contract management system checks it against the provider's policies and the governance rules.
3. **Secure model training.** Data aggregation and computation take place inside the Secure Data Environment. Raw data is not exposed to the consumer. Differential Privacy and immutable audit logs are used.
4. **Model validation and bias detection.** The trained model undergoes independent checks for bias, fairness, security, robustness, data leakage and explainability. Only models that pass proceed to pilot deployment and scaling.
5. **Deployment and continuous monitoring.** The model may be deployed for commercial or public use, subject to continuing monitoring, fairness and security logs, and traceability to the datasets and contracts used in training.

### Part C: Issues the draft leaves open

The proposal is ambitious, but several matters need legal and operational clarification. The third column, new in this edition, shows where AIGSI Version 3.0 already offers an answer that DEPA could adopt. Clause references are to the AIGSI Version 3.0 discussion paper.

| Issue | Question requiring clarification | Where AIGSI v3 offers an answer |
|---|---|---|
| **Legal basis** | What legal basis permits the original collection, and the subsequent processing or sharing, of personal data for AI training? | Part A: every obligation traced to DPDPA or ITA 2000. AIGSI does not create a legal basis; it requires one to be recorded. |
| **Anonymisation** | Who decides that data has been sufficiently anonymised, and who is responsible if re-identification becomes possible? | B.6: whoever supplied the data answers for its use. P2: unknown risk is treated as significant risk. |
| **Data ownership** | What rights do data providers, Data Principals and Data Collabs have over contributed datasets? | Copyright control area (D.I); LC-14 IP indemnity. Ownership itself is a policy question AIGSI leaves to contract and law. |
| **Commercial rights** | Who owns the resulting model, and how are intellectual property and commercial benefits shared? | MO-9: whoever fine-tunes, quantises or distils a model is the Model Owner of the modified model. LC-14 IP indemnity. |
| **Accountability** | How are legal responsibilities divided among the provider, the collective, the platform, the processor and the AI developer? | Four roles (B.1), backward liability chain, B.6 rules where roles blur, B.7 failure allocation. |
| **Model leakage** | Who is responsible if a trained model reveals personal or confidential information despite the safeguards? | B.7 failure allocation; release controls and containment testing (D.I); F.8 post-incident review. |
| **Cross-border use** | What restrictions apply when a model, an infrastructure provider or a consumer operates outside India? | B.5: recourse stops at a supplier beyond effective jurisdiction; DS-44 residual loss cover; LC-12 exit plan. |
| **Oversight** | How does the proposed consortium relate to existing statutory regulators and authorities? | F.0: one audit, three opinions (CERT-In, DPDPA, IT Act); DS-38 reporting matrix keeps statutory duties separate. |

The document's reference to de-identifying personal data before sharing needs particular attention. Pseudonymisation and anonymisation are not interchangeable. Pseudonymised information remains personal data wherever an individual can still be identified using additional information.

The legal consequences must be examined under the applicable law, and cannot be assumed away by the technical architecture.

## Part D: Issues for Data Fiduciaries and Independent Data Auditors

When a Data Fiduciary adopts this framework, it creates a new challenge for the Independent Data Auditors who assess DPDPA compliance. The auditor must be satisfied that personal data exported by the Data Fiduciary to the DEPA-AIT network does not create harm or risk for the Data Principals concerned.

Two fundamental questions arise:

1. Does the ability to share data securely automatically create the legal authority to share and process it?
2. Does making Indian datasets available through this architecture strengthen India's AI sovereignty, or could it let foreign-controlled entities capture the strategic and economic value of Indian data? This is a question of ethics and responsible use of AI.

The concerns below follow from these questions. Where AIGSI Version 3.0 has a clause that helps, it is noted under the heading "AIGSI v3".

### 1. Data availability is not authority to use data

DEPA-AIT brings TDPs and TDCs into a common techno-legal ecosystem. But under DPDPA, processing of personal data must rest on a ground recognised under Section 4, read with Sections 5, 6 and 7 as applicable. Being a registered DEPA participant does not create an additional ground. DPDPA recognises no general ground called "AI innovation", and gives no blanket permission to use personal data for AI training because the training may eventually benefit society.

The Data Fiduciary must first decide whether the proposed processing is lawful. That means answering questions such as:

- Was the data collected for a purpose that covers the proposed AI training?
- Does the original notice adequately describe that purpose?
- Is consent required, and if so, has valid consent been obtained?
- Can a specific legitimate use under Section 7 be established?
- Is any statutory exemption actually available?
- Do other laws, contractual obligations or sectoral regulations impose restrictions?

A data exchange platform cannot answer these questions. DEPA can facilitate access to data. It cannot manufacture a lawful basis for processing it.

### 2. Confusing pseudonymisation with anonymisation

The document rightly recognises that removing direct identifiers such as names, addresses and identification numbers does not prevent re-identification, because datasets can be linked and individuals identified through combinations of innocuous attributes. It therefore proposes more sophisticated techniques, particularly Differential Privacy.

Yet the same document requires de-identification with "pseudonymized keys held by the originating agency". If the originating agency keeps keys that can link the data back to an individual, can the

shared information automatically be treated as non-personal? It cannot. Pseudonymisation is a security technique. It is not anonymisation in the legal sense.

The document admits that combining datasets raises the risk of identification, and still relies heavily on the data provider's own responsibility to anonymise before sharing. That is not enough. A Data Fiduciary cannot discharge its statutory duty merely by declaring its dataset anonymised. It needs a defensible assessment of identification risk, covering linkage with other datasets, the nature of the information, the technical capability of likely recipients and the consequences of re-identification.

If a standardised DEPA certification of anonymisation is introduced, it must not become an automatic legal safe harbour.

> **AIGSI v3.** Principle P2 treats unknown risk as significant risk. An anonymisation claim that has not been tested against linkage is therefore an unknown risk, to be handled as significant. Under B.6, whoever supplied the data answers for its use, so the TDP cannot pass the anonymisation question to the platform.

### 3. Differential Privacy is no substitute for lawful processing

The proposal gives great weight to Differential Privacy, which is a valuable safeguard. But three separate questions must be kept apart:

1. Is the original collection and processing of personal data lawful?
2. Is the transformation and use of the data sufficiently privacy-preserving?
3. Does the resulting model or output create risks for individuals?

Differential Privacy mainly addresses the second and third questions. It does not answer the first. And a statement that Differential Privacy has been used does not by itself prove that privacy risk is controlled. The implementation must be independently evaluated.

### 4. Purpose limitation: can data collected for one purpose serve another?

This is perhaps the most serious DPDPA issue for an individual Data Fiduciary.

Section 5 and the DPDP Rules 2025 require a notice describing the personal data and the purposes of processing. A broad purpose such as "improving services" cannot be read as authorising every future AI training activity. The Data Fiduciary must examine whether the new purpose is covered by the original notice and lawful basis, or whether a fresh basis and a new notice or consent are needed.

AI training is also rarely a single, clearly defined activity. One dataset may be used to train several models, fine-tune existing models, validate algorithms, test competing products and build commercial applications. A general permission to contribute data to a training platform could thus open the door to many downstream purposes never explained to the Data Principal.

DEPA must not become a means for a Data Fiduciary to turn a narrowly collected dataset into a reusable commercial resource without checking the legal basis for each processing activity.

### 5. Consent: who obtains it, for what, and from whom?

The document introduces Data Collabs as intermediaries that may work with communities to aggregate, standardise and organise data, and that "may be responsible for obtaining primary consent where applicable". This raises several questions:

- If a patient consents to a hospital for treatment, can a Data Collab later treat that consent as permission for AI training?
- If a Data Collab obtains consent, is it acting as the hospital's agent, as a Data Processor, or as an independent Data Fiduciary?
- Who gives the notice required under DPDPA?
- Who enables withdrawal of consent?
- What happens when consent is withdrawn after the data is already in a training dataset?

These go to the heart of the DPDPA consent architecture. Consent under the Act must be free, specific, informed, unconditional and unambiguous, given by a clear affirmative action, and specific to a purpose. Consent taken once at collection cannot become a permanent authorisation for an open-ended range of AI applications. A person may agree to her health data being used for research into a particular disease, but not to its use in an unrelated commercial profiling system.

Withdrawal matters equally. A framework that can take data into a training environment but cannot respond meaningfully to withdrawal tilts the balance from the individual towards the convenience of AI developers.

The research exemption under Section 17(2)(b) also needs care. It is subject to statutory conditions and prescribed standards, and cannot be read as a blanket exemption for all commercial AI training.

### 6. Erasure after AI training

DEPA envisages a lifecycle in which data is contributed, transformed, used for training, validated and finally deployed through a model. What happens when a Data Principal exercises the right to erasure under Section 12?

Deleting the source record may be possible, and so may deleting the copy on a Data Exchange Platform. But what about:

- the transformed training dataset
- intermediate training files
- checkpoints and model weights
- embeddings and derived representations
- validation datasets
- backup copies
- models already deployed to customers?

Can the model be shown not to retain or reveal the individual's data? Can the contribution be removed? Must the model be retrained, fine-tuned or put through machine unlearning? Who pays for it?

DPDPA prescribes no universal technical method for removing an individual's contribution from a trained model, and erasing the source record does not erase its influence on the model. DEPA must solve this before any data is contributed. A Data Fiduciary should not enter a data-sharing arrangement without understanding what withdrawal, erasure, retention and model reuse will mean.

> **AIGSI v3.** The wider Model Owner definition (MO-9) helps answer "who pays". A TDC that fine-tunes, quantises or distils a model on DEPA data becomes the Model Owner of the modified model, and carries the Model Owner duties for it. The AI Bill of Materials required where models are routed or embedded (B.6) gives the traceability from model back to dataset that an erasure request needs.

### 7. A Secure Data Environment is not a legal sanctuary

The document relies heavily on SDEs: hardware-based secure enclaves, encryption, access controls and immutable logs. These can greatly reduce the risk of unauthorised access and leakage. They do not make every activity inside them lawful.

An unauthorised purpose stays unauthorised inside a secure enclave. A model trained on personal data without a valid legal basis does not become lawful because the data never left the enclave. Nor can a secure environment cure inappropriate data selection, excessive collection, unauthorised secondary use, inaccurate data, discriminatory training data, improper retention or unlawful disclosure of model outputs.

The distinction is simple. Security controls protect the processing environment. Data protection law decides whether the processing should happen at all, and on what conditions. The framework should say so explicitly.

> **AIGSI v3.** This distinction is built into the AIGSI audit. Under F.0 one engagement gives three separate opinions: a CERT-In assurance opinion on the security of the AI system, a DPDPA opinion on personal data, and an IT Act opinion on the security of non-personal data and systems. A clean security opinion on the SDE cannot be read as a DPDPA opinion on the processing.

### 8. The Data Fiduciary cannot outsource accountability to DEPA

The ecosystem involves Data Principals, Data Collabs, TDPs, TDCs, Data Exchange Platforms, SDE providers, technology service providers and an oversight body. That is a long chain of responsibility, and DPDPA does not recognise most of these as statutory categories.

The Act knows only Data Fiduciaries, which determine the purpose and means of processing, and Data Processors, which process on a fiduciary's behalf. A participant does not become a Data Processor because the framework calls it a service provider. An AI developer that receives data and independently decides how to train a commercial model is unlikely to be merely a processor for the originating institution. Depending on the arrangement, it may be a Data Fiduciary in its own right. Likewise, a Data Collab that aggregates data from several institutions and decides how it will be used may carry responsibilities that no contractual label can settle.

Section 8(1) makes the Data Fiduciary responsible for processing done by it or on its behalf by a Data Processor, and Section 8(2) requires a valid contract to engage a processor. An electronic contract

saying that the platform, exchange or AI developer will take care of compliance cannot remove that statutory responsibility. A contract can allocate operational duties and provide indemnities. It cannot override the statute.

For each processing activity, the framework must therefore identify:

- who determines the purpose
- who determines the essential means
- who acts on behalf of another entity
- who is an independent Data Fiduciary
- who responds to Data Principal rights
- who handles breach notification
- who decides on erasure and retention

Without this allocation, the framework risks creating an accountability vacuum.

> **AIGSI v3.** AIGSI's rules for blurred roles (B.6) apply directly to DEPA: a party that fine-tunes a model becomes its Model Owner; an integrator that recommends a model or host answers for that choice; where one organisation holds several roles, an internal agreement names an owner and a Human Handler for each. B.7 then gives guidance on which party should answer for which kind of failure, for use in drafting indemnities. None of this changes the Data Fiduciary's statutory accountability; it fills in the chain behind it.

### 9. Data sovereignty: a gateway for foreign control of Indian data?

The document speaks of India's ambition to build its own AI ecosystem, yet the architecture is not reserved for Indian AI developers. TDCs may be start-ups, researchers or enterprises, and the framework openly envisages global collaboration, cross-border research and international data-sharing arrangements.

International collaboration is welcome in science and technology, and Indian institutions should take part in global research. But sharing data on terms that preserve Indian control is very different from an architecture in which foreign entities may capture the economic and strategic benefit of Indian data without matching Indian control.

The draft sets no Indian ownership or nationality requirement for TDCs. Nor does it provide any framework for keeping Indian control over the resulting models, derived intelligence and commercial value. This is a significant omission.

An institution created to serve a national data need can, over time, pass into the control of a global information company. The law may regulate its operations, access to information and security. But who controls the company, who owns the technology, who builds the analytics and who profits from the intelligence are separate questions. CIBIL showed that an Indian regulatory framework does not by itself guarantee Indian ownership or strategic control over an institution handling nationally important information. That experience should be treated as a warning while designing DEPA.

### Sovereignty is more than localisation

Data sovereignty cannot be reduced to whether the server sits in India. Localisation helps with some jurisdictional and security concerns, but it does not keep national control over the data ecosystem. There are at least four dimensions:

1. **Physical control:** where the original dataset is stored and processed.
2. **Access control:** who can access the dataset and authorise its use.
3. **Intellectual property control:** who owns the model, algorithms and derived IP created from the dataset.
4. **Economic control:** who receives the commercial benefit of the intelligence extracted.

A framework that protects only the first two may still let the strategic value of Indian data migrate to foreign-controlled entities. This matters most in healthcare, financial services, agriculture, education, geospatial information and public administration. Such datasets reveal India's population, consumption patterns, disease distribution, credit behaviour, agricultural productivity, geographical vulnerabilities and economic activity. Combined and analysed at scale, they become a strategic national resource.

### DPDPA alone does not secure sovereignty

DPDPA is a personal data protection law, not a national data sovereignty law. Section 16 allows the Central Government to restrict transfers to notified countries or territories, alongside other legal restrictions. It does not require personal data to stay in India.

A Data Fiduciary should not assume that DPDPA compliance keeps the economic and strategic value of Indian data under Indian control. Some datasets will fall outside DPDPA altogether, because they are genuinely anonymised or are not personal data, but their strategic importance remains. That is a policy gap. DEPA must therefore deal with personal data protection and with the wider question of national data governance.

### Where the liability chain ends

AIGSI Version 3.0 adds a point that sharpens the sovereignty concern. Its new section B.5 records where backward recourse actually stops. Two of its cases describe exactly what DEPA could produce:

- **A supplier beyond effective jurisdiction.** If a foreign TDC misuses data or releases a leaking model, an Indian contributor's contractual remedy may be unenforceable in practice. Recourse then stops, in practice, at the Indian party.
- **An open-weight model under an as-is licence.** If a model trained on Indian data is published with open weights, recourse stops at whoever chose to use it. Nobody upstream answers for it.

In both cases the Indian TDP, and the Indian deployer that later uses the model, are left holding the risk while the value has moved abroad. AIGSI's answer is that a choice made without recourse is a liability decision, to be recorded and approved at the level that accepts the risk, with residual loss covered by insurance or reserves (DS-44) and an exit plan in the contract (LC-12). DEPA should require the same of every TDP before data leaves its control.

### Safeguards to build in

1. **Indian control over critical national datasets.** Data relating to banking, healthcare, public administration, national infrastructure and geospatial intelligence should carry clearly defined access and control conditions, not left to individual providers.
2. **Eligibility and scrutiny of foreign participants.** Foreign entities need not be excluded, but foreign-controlled participants should face scrutiny of ownership, beneficial ownership, jurisdiction and security, especially for strategic datasets.
3. **Ownership and control of trained models.** Every contract should state who owns the resulting models, weights, derived datasets, improvements and IP, and should address commercial exploitation outside India. Under AIGSI MO-9, a TDC that fine-tunes or distils a model on Indian data becomes its Model Owner; the contract should say what that ownership carries with it.
4. **Restrictions on onward transfer and affiliate access.** A foreign participant should not pass data, derived information, model artefacts or access rights to its parent, subsidiaries, affiliates or third parties without authorisation.
5. **Indian jurisdiction and enforceability.** Contracts should provide for enforceable Indian jurisdiction, audit access, regulatory cooperation, dispute resolution and remedies, backed by practical enforcement capability. Where enforceability is doubtful, residual loss should be covered (AIGSI B.5, DS-44, LC-13).
6. **Governance of Data Exchange Platforms.** Ownership and control of critical platforms should be disclosed, with safeguards against concentration of control and provision for portability, continuity, exit and independent audit.
7. **Benefit-sharing.** Where Indian datasets contribute materially to commercially valuable models, Indian institutions, contributors and the domestic AI ecosystem should receive a fair share of the value.
8. **Exit and change-of-control provisions.** A change in ownership, control or jurisdiction of a platform or a major TDC should trigger review of existing permissions and contracts.
9. **No open-weight release without approval.** Release of model weights trained on DEPA data should need the TDP's express approval, recorded as a liability decision under AIGSI B.5.

These safeguards still allow international collaboration. They ensure India does not give up control over strategically important data assets in the process.

## 10. Society-wide aggregation and collective profiling

The societal implications of DEPA may be graver than the compliance risk to individual institutions. The proposal aims to combine data across sectors, which is exactly what makes it attractive for AI, and exactly what enables profiling of individuals and communities in ways never contemplated at collection.

Combine banking transactions, health records, education records, geospatial information, insurance data and consumer behaviour. Even without direct identifiers, the result can reveal income, health, financial distress, social behaviour, location and employment. AI systems may learn correlations that no single dataset shows. Such models may then decide credit eligibility, insurance pricing, employment suitability or access to public services.

The individual may never know that data from a hospital, bank or government department fed a model that later decided something about her. De-identified training data does not remove the need to examine what the deployed model does. A national AI data ecosystem must include safeguards against discriminatory profiling, group harms, inappropriate inference and misuse of model outputs.

> **AIGSI v3.** Version 3.0 adds fairness testing and protection of vulnerable persons as deployer control areas (D.I), with matching upstream duties for developers. A deployer using a DEPA-trained model to decide credit, insurance or access to services would have to test it for group harm, not rely on DEPA's Step 4 validation alone.

### 11. Electronic contracts cannot replace statutory compliance

Machine-readable electronic contracts, specifying permitted purposes, security requirements and access conditions and enforced by technology, are a welcome innovation. But a contract between a TDP and a TDC is not a contract with every Data Principal whose data sits in the dataset. The Data Principal is not bound by it, and it cannot extinguish her statutory rights.

An electronic contract can record that the AI developer will not use data for an unauthorised purpose. It cannot prove that the original Data Fiduciary was entitled to share the data in the first place. It can require deletion after a set period, but it cannot settle statutory retention obligations or override a Data Principal's rights.

The framework should keep four layers of governance distinct:

1. the legal basis for processing personal data
2. the contractual conditions governing access and use
3. the technical controls that enforce those conditions
4. the accountability of each statutory Data Fiduciary

They must work together. None can replace another.

> **AIGSI v3.** AIGSI holds the same line. Under LC-1 the affected person's remedy against the deployer never depends on what recourse exists upstream. Under LC-5 liability caps do not limit statutory liability, and caps, exclusions and carve-outs must be disclosed before contract. New section A.5 maps the liability doctrines to DPDPA, ITA 2000, the Indian Contract Act and the Consumer Protection Act, so that contract recourse and statutory liability are kept apart.

### 12. The oversight body: a parallel compliance regime?

The document proposes an oversight body to set standards, register participants, oversee data-sharing procedures and facilitate dispute resolution, aligned with the IT Act 2000 and DPDPA 2023. What is its legal status? A statutory regulator, a consortium, an industry self-regulatory organisation, a government-authorised platform operator, or some mix?

DPDPA vests statutory functions in the Data Protection Board of India. A DEPA oversight body cannot assume or displace those powers. There is also a real risk of confusion if a participant treats a DEPA certification as proof of DPDPA compliance. Organisations would then face two assessments,

one under DEPA and one under DPDPA, and a good result in the first could breed unjustified confidence about the second.

The governance arrangement should state plainly that registration, certification, security audits and participation in DEPA confer no exemption or immunity under DPDPA.

> **AIGSI v3.** AIGSI's principle is that organisations should not be audited twice for the same system. Under F.0 an AIGSI audit is conducted as a CERT-In AI security audit, with AIGSI's DPDPA-centric and ITA 2000-centric parts forming the external-requirements layer of CERT-In's Statement of Applicability. Rather than build its own certification, DEPA could accept such an audit, with its three separate opinions, as the evidence it needs from participants.

### 13. Additional obligations for Significant Data Fiduciaries

This matters most for organisations classified as Significant Data Fiduciaries (SDFs): banks, hospitals, insurers, large platforms and others whose scale or manner of processing draws enhanced scrutiny. Section 10 requires SDFs to appoint a Data Protection Officer and an independent data auditor, among other measures, and the DPDP Rules 2025 add obligations on risk assessment and audit.

If an SDF contributes personal data to a DEPA training project, the decision cannot be treated as routine technology outsourcing. The SDF must assess the whole processing chain:

- the legal basis for the contribution
- the classification and sensitivity of the data
- the risk of re-identification
- the identity and statutory role of each participant
- the security and privacy controls of the SDE
- the possibility of model leakage
- contractual restrictions on downstream use
- the ability to respond to Data Principal rights
- the consequences of a breach or misuse
- ownership and control of the resulting models and derived intelligence
- where its recourse ends, and how residual loss is covered

The independent data auditor should examine the legal and governance basis for the SDF's participation, and not merely the platform's security certificate. An SDF cannot excuse a failure of its own obligations by pointing to the DEPA platform or an external provider's certification.

> **AIGSI v3.** The AIGSI audit has branches for Significant Data Fiduciaries and for RBI-regulated entities. Version 3.0 adopts CERT-In's grading of findings and remediation timelines (F.6A) and its rules on auditor competence and independence, with added legal competence for the DPDPA and IT Act parts (F.6B). These give the SDF's independent data auditor a recognised way to grade what it finds in a DEPA contribution.

## 14. What a Data Fiduciary should do before joining DEPA

DEPA should not be rejected. But no Data Fiduciary should join simply because it is presented as a national architecture for responsible AI. Before contributing any dataset containing personal data, the organisation should complete a structured legal and techno-legal assessment. I recommend these minimum checks:

1. **Data provenance.** Identify the source of every dataset, the purpose of collection, the notice given, the legal basis and any restrictions attached.
2. **Personal data classification.** Determine whether the dataset contains personal data, pseudonymised data, genuinely anonymised information, or a mix.
3. **Legal basis.** Establish the specific basis under Sections 4, 6 or 7, as applicable. Do not assume AI training is covered by the original purpose.
4. **Role determination.** Decide whether the organisation acts as a Data Fiduciary, a Data Processor or another participant, based on actual functions rather than contractual labels. Map each DEPA participant to its AIGSI role (Part E).
5. **Re-identification risk.** Assess the chance of identification through linkage, and the effectiveness of the proposed anonymisation and Differential Privacy.
6. **Contracts.** Ensure agreements cover permitted purposes, prohibited uses, retention, onward-sharing limits, security obligations, audit rights, incident reporting and liability, including disclosed caps and exclusions, insurance and IP indemnity (AIGSI LC-5, LC-13, LC-14).
7. **Data Principal rights.** Establish how withdrawal of consent, access, correction, erasure and grievance redressal will work across the data-sharing and training lifecycle.
8. **Model risk.** Evaluate the risk of leakage, memorisation, inference, discriminatory outcomes and other harms from the trained model, and the release controls that apply before a model leaves the SDE.
9. **Security.** Evaluate the SDE's encryption, access controls, privileged access, logging, monitoring, vulnerability management and incident response.
10. **Sovereignty and value retention.** Determine who controls the platform, who owns the model and its improvements, whether foreign affiliates can reach derived information, and whether the Indian contributor keeps appropriate rights and benefits.
11. **Where recourse ends.** Record, for each downstream party, where contractual recourse stops (AIGSI B.5), and how residual loss will be covered (DS-44). Have the decision approved at the level that accepts the risk.
12. **Independent audit.** Obtain an independent assessment of legal, technical and organisational safeguards. For an SDF, integrate it with the statutory audit.

This assessment should be complete before the first dataset is transferred or made accessible.

## Part E: How AIGSI Version 3.0 can fill the gap

DEPA is concerned mainly with enabling data access and secure AI training. It is not a complete compliance framework for the Data Fiduciary, which is why an integrated approach to data governance and AI governance is needed.

### What AIGSI is, and the status of Version 3.0

AIGSI (Artificial Intelligence Governance Standard of India) is FDPPI's AI governance standard and an enhancement of DGPSI-AI. Version 2.0 (8 November 2026) unified AIGSI with the RBI Guidance on Model Risk Management, CERT-In's guidelines on AI-accelerated vulnerability protection and on AI security auditing, and the DSCI paper on AI liability. It anchored every obligation in DPDPA or ITA 2000, set liability to flow backwards from the deployer to the developer and up to the model owner through a contract at each step, and split the audit into a DPDPA-centric part and an ITA 2000-centric part with a common core.

Version 3.0 has been prepared as a discussion paper for the Special Interest Group on AI Governance that FDPPI is forming. It is a draft for discussion, and Version 2.0 remains the reference text until the SIG concludes. Version 3.0 is backward-compatible: every earlier clause keeps its number.

### The relationship between the instruments

DEPA-AIT and AIGSI work at different layers. DEPA is a data-access architecture. It decides how data moves from a Training Data Provider into a Secure Data Environment and how a model comes out. AIGSI is an organisational governance standard. It decides whether each participant has governed its own part of that journey, and who answers for it when something goes wrong.

| Instrument | Role |
|---|---|
| **DEPA** | Enables data sharing for AI training |
| **DPDPA** | Governs lawful processing of personal data |
| **ITA 2000** | Governs security of electronic records and systems, including non-personal data |
| **DGPSI-AI** | Supports organisational governance of AI-related processing |
| **AIGSI** | Addresses the wider governance of AI systems across the lifecycle, and the liability chain behind them |

Each has its own function, and none can stand in for another.

### Mapping DEPA participants to AIGSI roles

AIGSI recognises four roles: Deployer, Developer, Model Owner and Infrastructure Provider. Version 3.0 divides the Developer into three sub-roles (product companies, tool providers and integrators) and widens the Model Owner to include anyone who fine-tunes, quantises, distils or republishes a model. DEPA's participants map onto these roles as follows.

| DEPA participant | AIGSI role | What AIGSI Version 3.0 asks of it |
|---|---|---|
| **Training Data Provider (bank, hospital, public agency)** | Data supplier; Data Fiduciary under DPDPA | Lawful basis and purpose check before contribution; DPDPA-centric audit of outgoing data; answers for the use of the data it supplied (B.6); records where its recourse ends (B.5) |
| **Data Collab** | Integrator (Developer sub-role) or Data Fiduciary, by function | Answers for the choices it recommends (B.6, DV-39); clear statutory role; internal owner and Human Handler for each role it holds |
| **Data Exchange Platform and SDE provider** | Infrastructure Provider | ITA 2000-centric security audit against CERT-In's nine domains (F.7); security operations; upstream duties including IP-6 |
| **Training Data Consumer that trains or fine-tunes a model** | Developer, and Model Owner of the model it modifies (MO-9) | Model risk management; leakage and bias testing; release controls; Model Owner duties MO-1 to MO-9; disclosed caps, insurance and IP indemnity (LC-5, LC-13, LC-14) |
| **Foundation model provider whose model is fine-tuned** | Model Owner of the base model | End point of the liability chain; but B.5 records where recourse stops in practice (as-is licence, capped terms, jurisdiction) |
| **Bank, insurer or agency that deploys the trained model** | Deployer | Accountable to the affected person regardless of upstream recourse (LC-1); fairness testing, vulnerable persons, residual loss cover (D.I, DS-44); kill switch and Human Handler |

### Six gaps in DEPA-AIT that AIGSI Version 3.0 addresses

#### 1. The liability chain, and where it ends

DEPA has a chain of participants (TDP → Data Collab → DEP/SDE provider → TDC → model provider → deployer) but allocates no liability along it. AIGSI sets liability to flow backwards through a contract at each step. Version 3.0 adds what Version 2.0 did not say: that the chain often ends before it reaches the party at fault. B.5 lists the situations (as-is licences, capped standard terms, private deployment, insolvency, a supplier beyond jurisdiction). DS-44 and LC-13 require residual loss to be covered. For DEPA, where many TDCs will be foreign and many models may be released openly, this is the most important single addition.

#### 2. Who is who when roles blur

DEPA defines its participants by function in the network, not by legal role. AIGSI's B.6 supplies the rules DEPA lacks. A TDC that fine-tunes a model on DEPA data is the Model Owner of the modified model, and whoever supplied the data answers for its use. A Data Collab that recommends a model, host or architecture answers for that choice; if a client mandates a choice against its advice, its duty narrows to a written warning. A party that routes work across several models answers for evaluating

each of them and lists them in the AI Bill of Materials. B.7 then offers guidance on which party should answer for which kind of failure, for use in drafting indemnities.

### 3. Personal and non-personal data

AIGSI's audit has a DPDPA-centric part covering personal data and an ITA 2000-centric part covering the security of non-personal data and of the AI system itself. This matters because much of DEPA rests on the claim that data has been "anonymised". Data that leaves the scope of DPDPA does not leave the scope of governance: the ITA 2000-centric audit still covers it, together with the sovereignty concerns in point 9 of Part D.

### 4. Model validation and release

DEPA's Step 4 ("fit for use" validation) has no defined standard. AIGSI's model risk management requirements, drawn from the RBI Guidance, provide one: model inventory, tiering, validation, leakage and bias testing, and monitoring after deployment. Version 3.0 adds release controls and containment testing as control areas, and fairness testing for deployers. Together these answer the question of what a model must pass before it leaves the SDE and who signs off its release. New principle P17, "prevent first, halt last", sets the order: prevent harm through release controls; keep the independent kill switch (P9) as the last resort. P17 also warns that over-control is a risk of its own, which matters for a framework meant to enable innovation.

### 5. One audit, three opinions

DEPA requires SDE providers to be "registered" but sets no audit standard, and its proposed oversight body risks creating a second compliance regime. Version 3.0 aligns AIGSI with the CERT-In AI Security Auditing Guidelines. Under F.0 one engagement produces a CERT-In assurance opinion, a DPDPA opinion and an IT Act opinion. F.7 maps CERT-In's nine audit domains to AIGSI clauses, giving the auditor a benchmark for the SDE and the training pipeline. F.6A and F.6B adopt CERT-In's grading of findings, remediation timelines, and auditor competence and independence rules.

### 6. Copyright, synthetic content and incidents

DEPA is silent on the copyright status of training data and model outputs, and on what happens after an incident. Version 3.0 adds copyright and synthetic media as control areas, an IP indemnity clause (LC-14), and a post-incident review procedure (F.8) that turns DSCI's 21 interrogation lenses into questions mapped to AIGSI clauses. A model that leaks personal data after leaving the SDE would be reviewed through that procedure, and the reporting matrix (DS-38) keeps DPDPA, CERT-In and sectoral reporting as separate duties.

### Alignment with MeitY's own guidelines

Part G.2 of AIGSI Version 3.0 maps the standard to MeitY's India AI Governance Guidelines (2025), as well as to ISO/IEC 42001 and the NIST AI RMF. MeitY's DEPA handbook and MeitY's AI governance guidelines should point in the same direction. An organisation that follows AIGSI can show alignment with both.

### In short

A DEPA participant that also adopts AIGSI can show an auditor, a regulator or the Data Protection Board that it has done more than use a secure platform: it has governed its own role in the data-to-model lifecycle, it knows where its recourse ends, and it has covered the gap. DEPA should therefore recognise such organisational standards explicitly, and not treat its own registration or certification as a substitute for them.

Professionals who wish to help refine AIGSI Version 3.0, including its application to frameworks such as DEPA-AIT, are invited to join FDPPI's Special Interest Group on AI Governance.

## Part F: What MeitY should correct before finalising the framework

The final version should carry explicit safeguards for both statutory compliance and national data sovereignty. I recommend that it state this principle:

> Participation in DEPA shall not, by itself, constitute a lawful basis for processing personal data under the DPDPA, nor shall registration, certification or use of a Secure Data Environment relieve any Data Fiduciary of its statutory obligations.

Every Training Data Provider should also be required to document:

1. whether the dataset contains personal data
2. the legal basis for the proposed processing
3. the purpose for which the data is being made available
4. the assessment supporting any claim of anonymisation
5. the identity and legal role of each participant
6. the safeguards for responding to Data Principal rights
7. responsibility for breach notification and remediation
8. arrangements for retention, deletion and model-level risk management
9. where contractual recourse against each downstream party ends, and how residual loss is covered

In addition, the architecture should include a separate data sovereignty policy covering:

- ownership and beneficial ownership of critical platform operators
- eligibility and scrutiny of foreign-controlled participants
- control over access to strategically important Indian datasets
- ownership and commercial use of trained models and derived intelligence
- restrictions on onward transfer to foreign affiliates
- Indian jurisdiction and enforceability
- change-of-control and exit arrangements
- benefit-sharing with Indian data contributors and the domestic AI ecosystem
- approval by the data provider before weights of a model trained on its data are released openly

Drawing on AIGSI Version 3.0, the framework should also:

- **define liability along the chain**, assigning each DEPA participant a role (Deployer, Developer, Model Owner or Infrastructure Provider) and requiring contracts that disclose caps and exclusions and provide for insurance and IP indemnity
- **treat a TDC that fine-tunes or distils a model as its Model Owner**, with the duties that go with it
- **set release controls** that a model must pass before it leaves the SDE, with a named person who signs off the release

- **accept one independent audit** producing separate security, DPDPA and IT Act opinions, instead of creating a parallel DEPA certification
- **require fairness testing and protection of vulnerable persons** before a DEPA-trained model is used to make decisions about individuals
- **recognise organisational governance standards** such as AIGSI explicitly

International collaboration should remain possible, on transparent conditions that protect individuals, Indian institutions and the national interest.

### Conclusion: India must not become a mere supplier of training data

India needs a national architecture for AI training, and nobody disputes that. Indian AI models built on Indian datasets can bring enormous economic and social value to healthcare, banking, education, agriculture and public administration.

But the drive to make India data-rich for AI must not end with individuals losing control over the lawful use of their personal data, or Indian institutions losing control over the value generated from the data they contribute. DPDPA was enacted to balance both: the individual's right to protect personal data, and the need to process it for lawful purposes.

DEPA has made a useful contribution by identifying technical mechanisms for sharing and processing data with reduced exposure. Its success should be judged by more than the volume of data released, the number of models trained or the value generated. The test must also include whether individual rights are protected, whether each Data Fiduciary can demonstrate compliance, and whether India retains real control over the strategic value of its data.

CIBIL reminds us that an Indian institution holding nationally important information can pass into foreign corporate control. The AI era multiplies the scope for value extraction, because the benefit now lies in the underlying data and also in the models, algorithms and intelligence built from it. AIGSI Version 3.0 adds a further warning: when something goes wrong, the chain of recourse may stop at the Indian border, leaving Indian institutions and citizens to bear the loss.

We face two dangers. One is that the convenience of a national data-sharing infrastructure becomes an excuse to weaken purpose limitation, consent, accountability and individual rights. The other, equally serious, is that India supplies the raw material for global AI while ownership and profits accumulate elsewhere.

A Secure Data Environment may stop a hacker from stealing a database. Differential Privacy may reduce the chance of identifying an individual from model outputs. An electronic contract may restrain a recipient from misusing data. None of them answers the fundamental legal question:

***Was the Data Fiduciary entitled to use this individual's personal data for this purpose in the first place?***

And none answers the equally important national question:

***Who will own, control and benefit from the intelligence created from Indian data, and who will answer for it when it causes harm?***

These questions must sit at the centre of India's AI data governance architecture. Otherwise India may supply the data, foreign entities may own the intelligence, and Indian society may end up paying for access to AI capabilities built from its own data.

Data empowerment without data accountability is not data protection. Data sharing without data sovereignty may become data extraction.

**Naavi**

## Annex: Questions for Discussion

Professionals reviewing this paper are invited to answer the questions below. Please quote the question number, say whether you agree with the position taken in the paper, and give your reasons. Partial responses are welcome. Responses may be sent to naavi@fdppi.in with the subject line "DEPA for AI Training".

Questions 1 to 35 are unchanged from the first edition. Questions 36 to 41 are new, and Question 29 has been updated to refer to AIGSI Version 3.0.

### A. Policy and purpose

1. Is the "Data Deadlock" a genuine problem that needs a new national architecture, or can it be solved within the existing consent framework of DPDPA?
2. Should access to DEPA-AIT be limited to Indian-owned and Indian-controlled entities? If foreign participants are allowed, on what conditions?
3. Should Government-held datasets (for example Nikshay, e-Sanjeevani, GST filings, Academic Bank of Credits, genomic data) be made available through DEPA-AIT at all? Which categories should be excluded outright?
4. Should datasets created with public money carry a royalty or benefit-sharing obligation? How should their value be determined, in line with the CAG's observations on valuing data?
5. Should DEPA-AIT remain a voluntary handbook, or be given statutory backing through rules under DPDPA or the IT Act 2000?

### B. Legal basis under DPDPA

6. What lawful basis can a Training Data Provider rely on to contribute personal data for AI training: consent, a legitimate use under Section 7, or the research exemption under Section 17(2)(b)?
7. Can the Section 17(2)(b) exemption cover commercial AI training? What conditions should the Rules prescribe for it?
8. Is pseudonymised data, where the originating agency keeps the keys, still personal data? Who should certify that a dataset is anonymised, and against what standard?
9. Can a model trained on personal data, even with Differential Privacy, itself contain personal data? Should model weights be treated as personal data for any purpose?
10. How should the rights to withdraw consent and to erasure (Sections 6(4) and 12) apply once data has been used to train a model?
11. Can the education use cases (adaptive learning, skills passports) be reconciled with the restrictions on children's data under Section 9?

### C. Roles and accountability

12. How should each DEPA participant (Training Data Provider, Data Collab, Data Exchange Platform, Secure Data Environment provider, Training Data Consumer, model provider) be classified under DPDPA: Data Fiduciary, Data Processor or otherwise?

13. If a Data Collab obtains consent on behalf of Data Principals, is it acting as a Consent Manager? Should it then register under Rule 4?
14. Who should be liable if a trained model leaks personal data or enables re-identification? Can an electronic contract shift that liability?
15. Should Data Collabs and Data Exchange Platforms be notified as Significant Data Fiduciaries, with DPIA and independent data audit obligations?
16. What should be the legal status of the proposed oversight body, and how should it relate to the Data Protection Board of India?

### D. Technical safeguards

17. Should downloading datasets (Option 1 of the handbook) be permitted for any data derived from personal data, or should all training be confined to Secure Data Environments?
18. Who should set the Differential Privacy budget (epsilon) for each sector, and should the values be published?
19. What tests should a model pass before it leaves the Secure Data Environment (for example extraction and membership-inference tests), and who should sign off on its release?
20. Should Secure Data Environment infrastructure be Indian-controlled, with hardware attestation verifiable in India?

### E. Data sovereignty

21. Who should own the model weights and intellectual property of models trained on Indian datasets?
22. Should foreign AI developers be allowed to fine-tune their foundation models on Indian datasets and deploy them globally? If so, on what terms?
23. Should onward transfer of data, derived information or model artefacts to foreign parent companies and affiliates be restricted?
24. Should a change in ownership or control of a platform or a major Training Data Consumer automatically trigger a review of its access, given the CIBIL experience?
25. Who should negotiate the cross-border data agreements envisaged in the roadmap: the Central Government alone, or platform operators as well?
26. Is a model of accredited “Indian Data Training Companies”, acting as a protective layer between data providers and AI developers, a workable alternative?

### F. Data Fiduciaries and Independent Data Auditors

27. What should an Independent Data Auditor verify before accepting a Data Fiduciary’s decision to contribute data to DEPA-AIT?
28. Should DEPA registration or certification be accepted as evidence of “reasonable security safeguards” under Section 8(5)?
29. How should standards such as DGPSI-AI and AIGSI Version 3.0 be applied to DEPA participants? Should DEPA recognise such organisational standards formally? *(updated)*

30. What additional checks should a Significant Data Fiduciary, such as a bank or hospital, complete before joining DEPA-AIT?

### G. Society and the Data Principal

31. What safeguards are needed against collective profiling and group harms when data from banking, health, education and geospatial sources is combined?
32. Should Data Principals have a public registry of datasets and consumers, a collective opt-out, and a grievance channel within DEPA?
33. Does DEPA-AIT, as drafted, risk being seen as a dilution of DPDPA, and could it affect the Government's defence of the Act before the Supreme Court?

### H. Overall view

34. On balance, should MeitY adopt DEPA-AIT as drafted, adopt it with modifications, or withdraw it for a fresh consultation?
35. What are the three most important changes you would recommend to MeitY?

### I. AIGSI Version 3.0 and DEPA-AIT (new)

36. Is the mapping of DEPA participants to AIGSI's four roles (Deployer, Developer, Model Owner, Infrastructure Provider) in Part E correct? Which participant is hardest to classify?
37. Should a Training Data Consumer that fine-tunes or distils a foundation model on Indian data be treated as the Model Owner of the modified model (AIGSI MO-9), with the duties that go with it?
38. Where a Training Data Consumer is beyond effective Indian jurisdiction, should the Training Data Provider be required to cover residual loss through insurance or reserves (AIGSI B.5, DS-44)? Who should bear that cost?
39. Should release of open weights of a model trained on DEPA data require the express approval of every Training Data Provider whose data was used?
40. Should DEPA accept a single independent audit producing separate CERT-In, DPDPA and IT Act opinions (AIGSI F.0) in place of its own certification of participants?
41. Would you be willing to take part in FDPPI's Special Interest Group on AI Governance, and which workstream interests you: liability and contracts, audit and assurance, technical controls, or legal and sectoral?

**Please complete this questionnaire online at here**