Foundation of Data Protection Professionals in India
[Not for Profit, Section 8 Company limited by guarantees]
CIN No: U72501KA2018NPL116325: GSTN 29AADCF4963H1ZC
Registered Office: No 37, "Ujvala", 20th Main, BSK first Stage, Second Block, Bangalore 560050
Web: www.fdppi.co.in : www.fdppi.in : E Mail fdppi@fdppi.in: Ph: 08026603490: Mob: +91 8310314516

AI Governance Standard of India

Version 1 — Draft for Discussion

About This Document

India currently has three separate, non-interoperable sources of AI governance obligation: a DPDPA-focused voluntary framework (DGPSI-AI), a prudential model-risk regime mandatory for RBI-Regulated Entities, and a cybersecurity/vulnerability-management regime aimed at OEMs and technology providers under CERT-In.

An organisation that deploys AI, is regulated by RBI, and also relies on third-party AI products must, in practice, satisfy all three at once — yet none of the three source documents is written with the others in view.

This Version 1 draft brings the three together into a single, structured standard with three parts:

- a set of Foundation Principles that generalise the values running through all three sources;

-a set of Deployer-Oriented Model Implementation Specifications, addressed to the organisation that puts an AI model into use (the "Deployer" — typically a Data Fiduciary under DPDPA and/or a Regulated Entity under the RBI Guidance); and

-a set of Developer-Oriented Model Implementation Specifications, addressed to whoever builds or supplies the model (the "Developer" — typically the Model Developer under the RBI Guidance and/or the OEM/Technology Provider under the CERT-In Guidelines).

This is a suggested framework to be developed over industry consultation into a more formal best practice standard.

Content

Section	Contents
Section 1 — Foundation Principles	12 principles covering accountability, risk-tiering, explainability, security-by-design, human oversight and lifecycle governance.
Section 2 — Deployer-Oriented Specifications	30 specifications, in 6 clusters, for the organisation putting an AI model to use.
Section 3 — Developer-Oriented Specifications	31 specifications, in 7 clusters, for whoever builds or supplies the model.

Definitions and Roles

The three source frameworks use overlapping but not identical vocabulary. This section fixes a common set of terms for the rest of this standard; where a source uses a different word for the same concept, that is noted.

Term	Definition
Model / AI Model	A system — whether developed internally, sourced from third parties, or a combination — that incorporates data, applies theoretical, empirical or judgement-based assumptions, and uses statistical, mathematical, economic, financial or other cognitive techniques (including AI/ML) to analyse, interpret and process inputs and produce outputs used for business or other decision-making, irrespective of whether the organisation itself labels the tool a "model".
Deployer	The organisation that puts an AI model to use in its business processes — typically a Data Fiduciary or Significant Data Fiduciary under DPDPA, and/or a Regulated Entity (RE) under the RBI Guidance. Referred to in DGPSI-AI simply as the "Deployer".
Developer	Whoever designs, develops, tests, trains and documents the model's methodology — the "Model Developer" under the RBI Guidance, or the OEM / Technology Provider (software vendor, hardware manufacturer, cloud service provider, managed service provider, system integrator) under the CERT-In Guidelines.
Human Handler	The specific, named individual accountable for the consequences of an AI model's use — on the deployer side, by default the DPO/Compliance Officer or process owner; on the developer side, disclosed as part of the licensing contract.
Explainability	The property of a model to express the important factors influencing its results in a way that is understandable to those relying on it.
Significant Data Fiduciary (SDF)	A Data Fiduciary notified as such under DPDPA, subject to enhanced obligations including DPIA, independent audit and a Data Protection Officer.
Regulated Entity (RE)	An entity to which the RBI Guidance applies — commercial banks, small finance banks, payments banks, regional rural and co-operative banks, NBFCs, All-India Financial Institutions, asset reconstruction companies and credit information companies.
Model Risk	The risk of adverse outcomes arising from model errors, misapplication of outputs, or a model becoming unfit/unsuitable over time.
Kill Switch	A tamper-proof mechanism, controlled independently of a model's own intelligence, by which an AI system can be forcibly deactivated; for humanoid/industrial robots, controlled on a separate chip/circuit from the device's intelligence.
Critical Risk AI	An AI system classified as posing critical risk on account of its autonomy, the sensitivity of data it processes, and its automated decision-making capability's potential effect on society — includes cyborgs and sentient algorithms by default.

Section 1 — Foundation Principles

DGPSI-AI's own six principles form the core of this section. Because DGPSI-AI addresses only DPDPA compliance, six further principles have been added — drawn from the RBI Guidance's model-risk governance and from CERT-In's security-by-design requirements — so that the combined set speaks to all three dimensions of AI governance: data protection, model risk, and cybersecurity.

No	Principle	Explanation
1	**Accountability Through a Human Handler**	Behind every AI algorithm there shall be one identifiable human accountable for its outcomes, on both the deployer and developer sides.
2	**Unknown Risk is Significant Risk**	Where an AI system's behaviour or risk cannot be fully known or explained, it shall by default be treated as a significant/high risk requiring enhanced governance.
3	**Explainability by Design**	Every privacy notice or disclosure covering an AI process shall be accompanied by an explainability disclosure; higher-impact models require higher explainability thresholds, and where full explainability is not achievable, compensating controls apply.
4	**Justified Use, Not Default Use**	The use of an AI process shall be validated by a document justifying the technical, operational and economic need, at both the deployer and developer ends, before it is put to use against a data principal.
5	**Guardrails Against Dark Patterns and Harm**	Every AI process shall document the specific guardrails securing the processing against dark patterns, neurological manipulation, and physical or psychological harm to any data principal.
6	**Fiduciary Responsibility to Society**	The deployer, as fiduciary, shall ensure all measures necessary to safeguard society at large from adverse effects arising from the use of AI — not the individual data principal alone.
7	**Risk-Based, Tiered Governance**	The intensity of governance — validation frequency, approval authority, monitoring and documentation — shall be proportionate to a model's materiality, complexity, autonomy and the degree of reliance placed on its output.
8	**Independent Validation and Challenge**	All models, including third-party and AI/ML models, shall be subject to independent validation or audit by a function distinct from the team that built, owns or commercially benefits from the model, following a three-lines-of-defence structure.
9	**Human Oversight and Override**	Human-in-command arrangements, override/suspension/kill-switch mechanisms, and periodic human review of AI-driven decisions shall be maintained, with explicit safeguards against automation bias and decision fatigue.
10	**Security and Integrity by Design**	AI systems and the infrastructure they run on shall be built and operated under a Secure Development Lifecycle, with continuous vulnerability assessment (including AI-assisted testing), access control and credential hygiene.
11	**Timely, Transparent Disclosure**	Vulnerabilities, security incidents and material model risks shall be disclosed to affected stakeholders and to the relevant regulator (Data Protection Board / CERT-In / RBI) within defined timeframes, without waiting for a complete post-mortem.
12	**Lifecycle Governance and Continual Improvement**	AI governance is not a one-time certification. Models shall be governed across their full lifecycle — selection, development, validation, approval, deployment, monitoring, change management and decommissioning — with records retained well beyond decommissioning.

Section 2 — Deployer-Oriented Model Implementation Specifications

These 30 specifications are addressed to the organisation that puts an AI model to use — the Deployer. They are grouped into six clusters. Deployers that are Significant Data Fiduciaries under DPDPA and/or Regulated Entities under the RBI Guidance are bound by the underlying source obligations already; deployers outside those categories can treat this section as a benchmark of good practice.

A. Governance and Accountability

No	Specification
DS-1	Designate a specific human handler — by default the DPO/Compliance Officer or the relevant process owner — accountable for the consequences of AI use in each process.
DS-2	Document the human handler for AI used under a vendor/licensing contract, and ensure it is recorded in the contract or cross-referenced where the developer has hardcoded an accountable person.
DS-3	Have the Board (or equivalent governing body) approve a Model/AI Risk Management Framework covering taxonomy, governance structure, risk tiering, inventory/documentation standards, and the full model lifecycle, delegating oversight to a Risk Management Committee.
DS-4	Ensure Senior Management allocates adequate human and technical resources to operationalise the framework, implement risk tiering, maintain the model inventory, and report periodically to the governing committee.

B. Risk Assessment and Classification

No	Specification
DS-5	Before deployment, obtain confirmation from the vendor on whether the software qualifies as "AI" (uses autonomous learning algorithms or probabilistic models producing outputs not fully predetermined by explicit code), and document a Risk Assessment treated as the DPIA for that process.
DS-6	Augment the DPIA/Risk Assessment with an independent external data auditor's evaluation at least once a year.
DS-7	Where prudent evaluation shows the "Unknown Risk" in a given process is not likely to cause significant harm, document an AI-Deviation Justification opting out of Significant-Data-Fiduciary-equivalent obligations for that process, with defensible reasoning on record.
DS-8	When assigning a risk tier to an AI model, additionally weigh the extent of reliance and level of autonomy placed on the model's output for decision-making, not materiality/complexity alone.
DS-9	For material third-party AI models, datasets and dependencies, separately assess supply-chain concentration risk, limits on independent validation, and the risk of provider-driven behavioural changes (e.g. silent model updates).

C. Vendor and Developer Due Diligence

No	Specification
DS-10	Collect an authenticated Explainability document from the developer, as part of the licensing contract, describing how the AI functions in personal-data processing and the likely harm it may cause data principals.
DS-11	Develop and retain an "AI Justification Document" before adopting an AI-led process for personal data under DPDPA, justifying the technical/economic need and the exposure of data principals to unknown risk.
DS-12	Obtain a documented assurance from the developer/licensor that the AI has been adequately tested for vulnerabilities (preferably by a third-party auditor), is reasonably secured against threats to confidentiality, integrity and availability of personal data, carries sufficient guardrails for data principals, and is free of malware.

No	Specification
DS-13	Prior to acquiring or using a third-party model, assess the credibility of the provider, the methodological soundness of the model and its limitations, and the suitability/quality of the data used — irrespective of any certification the provider offers.
DS-14	Ensure contracts for third-party AI models guarantee: minimum technical documentation sufficient to validate the model; audit rights for the deployer and its regulators, directly or through external experts; and continuity/exit arrangements.
DS-15	Where a vendor/OEM does not disclose adequate information about the AI model, identify the resulting risk and put in place mitigants such as restricting the model's use.
DS-16	As part of vendor onboarding and periodic renewal, request the deliverables set out in Section 3.G (Security Posture Assessment, Vulnerability Remediation Action Plan, Security Compliance Commitment, Continuous Assurance Report, SDL Compliance Certification).

D. Deployment Controls and Human Oversight

No	Specification
DS-17	Ensure deployment of an AI model does not introduce vulnerabilities into the production environment; implement access controls, cyber safeguards, and controls over external interfaces, APIs and integration pipelines.
DS-18	For AI models (including generative AI) interfacing with customers or external users: implement controls against prompt injection and adversarial inputs, limit session/context persistence, detect anomalous usage, disclose that users are interacting with an AI system (with its limitations), and offer an option to switch to human assistance.
DS-19	Establish human-in-command arrangements (human-in-the-loop / human-on-the-loop), override/suspension/kill-switch mechanisms, and periodic human review of AI-driven decisions to catch anomalies.
DS-20	Ensure oversight mechanisms explicitly guard against automation bias, over-reliance on model output and decision fatigue, and that personnel performing oversight have adequate expertise to challenge or escalate model outputs.
DS-21	Define explainability/transparency thresholds proportionate to a model's impact; where full explainability is not achievable, apply enhanced validation, testing, monitoring, usage restriction and other compensating controls.
DS-22	Put in place structured challenge processes (e.g. red-teaming) for models involving customer interaction or generative capability, and enhanced controls — justification, data-quality checks, more frequent monitoring — for models with dynamic or automatic updates.

E. Monitoring, Incident and Change Management

No	Specification
DS-23	Subject all deployed AI models to ongoing performance monitoring — backward- and forward-looking, including AI-specific evaluations and benchmarking — to ensure continued alignment with intended outcomes.
DS-24	Define what constitutes a material change to an AI model, requiring re-validation and re-approval, and maintain a comprehensive log of changes, versioning and approvals.
DS-25	Include AI-specific scenarios (unavailability, degraded or hallucinating output, provider withdrawal) in business-continuity planning, with defined fallback/manual-intervention arrangements.
DS-26	Establish an incident-management policy covering identification, validation and response to AI-related privacy/security incidents, and notify the Data Protection Board, CERT-In and other statutory authorities as required, without undue delay.

F. Documentation, Audit and Regulatory Engagement

No	Specification
DS-27	Maintain an accurate, up-to-date inventory of all AI models (active, inactive/under development, decommissioned) including owners, developers, validators, approvers, risk tier, intended use and dependencies; retain decommissioned models in the inventory for at least ten years.
DS-28	If the organisation is a Significant Data Fiduciary and/or a Regulated Entity, appoint an independent external data/model auditor with the necessary credentials and subject AI processes to periodic audit and DPIA at least annually.
DS-29	Promptly receive and respond to communications from the Data Protection Board, RBI (RMCB reporting) and/or CERT-In, participate in inquiry processes, and where available, avail of voluntary-undertaking provisions.
DS-30	Exercise verification rights over OEMs/technology providers — independent security assessments, vulnerability verification, patch validation, penetration testing, and requests for SDL/compliance documentation — as part of the deployer's own audit and vendor-oversight programme.

Section 3 — Developer-Oriented Model Implementation Specifications

These 31 specifications are addressed to whoever builds or supplies the model — the Developer, encompassing both DGPSI-AI's "AI developer" and CERT-In's "OEM / technology provider". They are grouped into seven clusters.

A. Explainability, Documentation and Risk Disclosure

No	Specification
DV-1	Generate an Explainability document describing the model's algorithmic function, using key principles of transparency, in a form the deployer can pass on to data principals.
DV-2	Provide the business contact details of the Human Handler of the AI model, responsible for its functioning, as part of the licensing contract.
DV-3	Document the training and testing process adopted in developing the model.
DV-4	Document a Risk Assessment of the model, including its susceptibility to third-party security compromise and its potential harm to users, data principals and society at large.
DV-5	Document the guardrails incorporated in the model to mitigate security risk, and the model's default configuration.
DV-6	Provide comprehensive instructions to users on any re-configuration or re-training suggested, required or expected in the model's normal use.
DV-7	Maintain and disclose to Indian customers and CERT-In an updated Bill of Materials — software, hardware, cryptography, AI and quantum — of deployed products, versions, exposed services, APIs and third-party dependencies.
DV-8	Document the use of AI agents as part of the developer's own workforce/build process, and their likely impact on the model developed.

B. Testing, Security and Secure Development

No	Specification
DV-9	Conduct comprehensive vulnerability assessments using both traditional security testing and AI-assisted vulnerability discovery (ML, LLMs, automated testing), covering source-code analysis, software composition analysis, dependency risk, threat modelling, penetration testing and behavioural anomaly detection.
DV-10	Assess and guardrail risks arising from the developer's own AI-enabled services, APIs, plugins and automation tools, based on data sensitivity, autonomy, connectivity and impact, with human oversight, monitoring/logging and dependency review as safeguards.
DV-11	Maintain evidence and periodic certification that AI-assisted code analysis, automated vulnerability discovery, security-token scanning, dependency analysis and advanced validation have been carried out using industry-recognised tools and methodologies.
DV-12	Implement and maintain a Secure Development Lifecycle across all stages of product/software development — secure architecture reviews, secure coding, threat modelling, source-code analysis, dynamic testing, penetration testing, dependency validation, software composition analysis, supply-chain risk management and secrets management — prior to release.
DV-13	Ensure products are free of hardcoded credentials, insecure default configurations, exposed administrative interfaces, unsupported third-party libraries, debug mechanisms or insecure authentication controls; validate all releases, patches, APIs and cloud components before release.

No	Specification
DV-14	Have the AI model audited by an independent third-party auditor using an acceptable audit standard, before and periodically after release.
DV-15	Ensure the model is not overfitted to training data, generalises appropriately (tested on out-of-sample data and varied scenarios), does not rely on spurious correlations, and that output variability/uncertainty under similar inputs is appropriately bounded and explained (confidence scores, probability outputs).
DV-16	Establish mitigants for data risk — quality, non-representativeness, incompleteness, IP breach — and monitor for data drift and concept drift on an ongoing basis, communicating material shifts to deployers.

C. Vulnerability Management and Patch Response

No	Specification
DV-17	Communicate any Critical (CVSS 9.0–10.0) or High (CVSS 7.0–8.9) severity vulnerability to affected Indian organisations and CERT-In immediately upon discovery or confirmation, with interim mitigation guidance and recommended remediation timelines.
DV-18	Upon discovering a zero-day vulnerability, active exploitation, or exploitation through AI-assisted techniques, notify affected organisations and CERT-In immediately, with interim safeguards, indicators of compromise, detection guidance and containment procedures, without delay.
DV-19	Establish accelerated patch management aligned to indicative timelines: Critical vulnerabilities discoverable via AI — emergency release (IT) / 7–15 days (OT); High — 7 days (IT) / 15–30 days (OT); responsibly-disclosed Critical — 5 days (IT) / 15–30 days (OT); responsibly-disclosed High — 15 days (IT) / 30–60 days (OT) — validated through appropriate testing and change management.
DV-20	Where immediate patching is not feasible, provide interim mitigation: virtual patching and compensatory controls, disabling vulnerable features, network segmentation, firewall/IPS rules, restricting internet exposure, enhanced logging, MFA enforcement, application allow-listing, and configuration hardening.
DV-21	Provide detailed patch-deployment documentation, testing guidance, rollback procedures, validation/integrity-verification mechanisms, compatibility considerations and technical support during deployment.
DV-22	Establish automated vulnerability/patch notification mechanisms to alert CERT-In and affected Indian organisations immediately when advisories, mitigations, patches or emergency measures become available.

D. Credential and Access Management

No	Specification
DV-23	Implement robust credential, identity and privileged-access controls across products, applications, support infrastructure, cloud environments and administrative interfaces — MFA, role-based access control, privileged access management, password rotation, just-in-time and time-bound administrative access, continuous authentication monitoring, and automated secret-scanning across code repositories — and avoid hardcoded credentials, embedded secrets or undocumented administrative access entirely.
DV-24	Conduct periodic credential-hygiene audits and maintain evidence that no customer-related credentials or secrets are exposed in public repositories, support systems or unauthorised environments.

E. Incident Response and Transparency

No	Specification
DV-25	Establish formal incident-response and vulnerability-disclosure processes for timely detection, containment, mitigation, recovery and customer coordination, communicated in writing to Indian customers and CERT-In.

No	Specification
DV-26	Upon discovery of active exploitation, compromise or a security incident, initiate response procedures, notify affected organisations, preserve logs/forensic evidence, identify indicators of compromise, assess impact, and provide containment, recovery guidance and coordinated remediation; report cyber incidents to CERT-In within 6 hours of notice, per Section 70B of the IT Act, 2000.
DV-27	Provide preliminary incident notification at the earliest possible stage (copying CERT-In), followed by detailed root-cause/forensic analysis and remediation status as soon as reasonably practicable, with regular status updates until resolution; retain relevant logs (firewall, VPN, authentication, endpoint, cloud, network, administrative activity) securely and in synchronised form to support investigation.

F. Safety Controls

No	Specification
DV-28	Incorporate a reasonably tamper-proof Kill Switch in the AI model, controlled on a separate chip/circuit independent of the model's own intelligence, configured to self-destruct if the model attempts to access it without human authorisation.
DV-29	Incorporate a set of emergency-handling instructions for scenarios where the AI risks hallucination or behaving outside its intended bounds.
DV-30	Where an AI is classified Critical Risk (by autonomy, sensitivity of data processed, or automated decision-making impact on society), make post-implementation behaviour monitoring available at the deployer's choice.

G. Deliverables to Deployers and Regulators

No	Specification
DV-31	Be prepared to provide, on request from an Indian deployer or CERT-In: (1) a Current Security Posture Assessment; (2) a Vulnerability Remediation Action Plan (CVE identifiers, CVSS ratings, affected versions, exploitation prerequisites, patch status, interim mitigations, testing/deployment timelines, rollback procedures, operational-impact considerations); (3) an Enhanced Security Compliance Commitment signed by senior management, naming a cybersecurity liaison officer; (4) a Continuous Security Assessment and Assurance Report (VAPT, breach-and-attack simulation, configuration reviews, independent audits, updated SBOM); and (5) an SDL Compliance Certification.

Vetted by Naavi