Ai-Assisted Personal Data Security Gap Assessment for Indian MSMES: Design and Implementation of a DPDP Rule 6 Decision-Support Prototype

Siddhanth D. Tarafder[1], Dr. Bonthu Kotaiah[2]
[1]*Department of Computer Science, Central Tribal University of Andhra Pradesh, Kondkarakam, Vizianagaram, India*
[2]*Associate Professor and Head, Department of Computer Science, Central Tribal, University of Andhra Pradesh, Vizianagaram, India*

***Abstract*—Indian Micro, Small and Medium Enterprises (MSMEs) may find it difficult to translate personal-data protection requirements into operational security controls, particularly when they have limited legal, technical and financial resources. This paper describes DPDP Sentinel, an academic decision-support prototype designed to help Indian MSMEs conduct a structured self-assessment of reasonable security safeguards associated with Section 8(5) of the Digital Personal Data Protection Act, 2023 and Rule 6 of the Digital Personal Data Protection Rules, 2025. The implementation combines a React interface, FastAPI backend, SQLite persistence, a deterministic assessment engine, a seven-category regulatory knowledge base, retrieval-supported explanations, optional Google Gemini generation, human-review flags, dashboard visualisations and printable report output. The repository contains 30 synthetic organisational scenarios and a module labelled as a comparison of rule-based, zero-shot LLM and hybrid approaches. Code inspection reveals, however, that the latter two outputs are simulated using programmed rules rather than generated by independently executed models; the reported metrics therefore do not yet establish valid comparative AI performance. This paper documents the implemented system, identifies methodological and field-level limitations, and proposes an evaluation plan using expert-labelled scenarios, actual model outputs and MSME usability testing. DPDP Sentinel is a research prototype, not a legal certification tool.**

***Index Terms*—Digital Personal Data Protection Act; DPDP Rules 2025; MSME; security safeguards; compliance decision support; retrieval-augmented generation; explainable AI.**

I. INTRODUCTION

Organisations that collect digital personal data must translate legal obligations into practices such as access governance, security monitoring, backup and recovery, retention management, processor oversight and staff procedures. Smaller organisations may lack dedicated privacy teams and may find consultant-led assessments costly or time-consuming.

This motivates an accessible, guided decision-support system that explains requirements in plain language, collects structured information about organisational practices, flags possible gaps and produces an actionable report. DPDP Sentinel is a working software prototype, not merely a conceptual framework. Its current scope is narrow: it focuses on Section 8(5) of the Digital Personal Data Protection Act, 2023 and seven safeguard categories encoded for Rule 6 of the 2025 Rules. It does not assess every obligation in the DPDP framework.

The paper's intended contributions are: (1) an implemented workflow mapping reported organisational practices to structured safeguard categories; (2) a decision-support design separating deterministic status assignment from natural-language explanation; and (3) a transparent plan for evaluating accuracy, source traceability, usability, time savings and affordability. Claims of real-world impact remain hypotheses until tested with participants.

II. REGULATORY AND RELATED BACKGROUND

A. Regulatory Context

Section 8(5) of the DPDP Act addresses reasonable security safeguards to prevent personal-data breaches. Rule 6 of the Digital Personal Data Protection Rules, 2025 specifies safeguards involving data security measures, access controls, logs and monitoring, continuity through backups, retention of relevant logs, processor-contract safeguards, and technical and organisational measures. The Rules have phased commencement provisions; the manuscript must state the applicable commencement status as of submission rather than implying that every provision became enforceable immediately upon notification.

R6-A through R6-G are internal project labels, not official statutory labels. The legal wording, applicability, commencement dates and any corrigenda must be checked against the official Act and Rules before submission.

B. Related Research

Kingston examined AI support for GDPR compliance through checklists, risk assessments, explanations and breach-related tasks, noting the value of rule-based approaches where legal justification and explainability matter. Lorè et al. presented INTREPID, an AI framework for analysing Italian public-administration documents for GDPR-related issues and security breaches using language processing, classification and expert annotation. Furthermore, recent legal analyses emphasize that informational privacy in the age of artificial intelligence requires robust regulatory frameworks, balancing automated decision-support accountability with transparent compliance verification. These works are relevant precedents but address different jurisdictions, populations and workflows.

The defensible research gap is not that no similar product exists. Rather, this project investigates an MSME-oriented, requirement-level self-assessment workflow grounded in India's DPDP Rule 6, with human-review flags and a planned evaluation of deterministic assessment plus source-grounded explanations. A systematic literature and product review is still needed before making a strong novelty claim.

III. PROBLEM STATEMENT AND RESEARCH QUESTIONS

Problem statement: MSMEs may need practical assistance translating personal-data security obligations into understandable controls and prioritised actions, but self-reported answers and AI-generated explanations must not be mistaken for proof of legal compliance.

Aim: To design and evaluate an accessible AI-assisted prototype that maps reported MSME security practices to structured Rule 6 safeguards and generates explainable findings for human verification.

- RQ1: Can a deterministic requirement-level engine produce consistent statuses from structured questionnaire responses?
- RQ2: Does retrieval-grounded explanation improve source traceability and expert-rated usefulness compared with an ungrounded LLM response?
- RQ3: Can the workflow reduce completion and review time compared with a baseline checklist?
- RQ4: What usability, privacy, trust and affordability conditions affect adoption by resource-constrained MSMEs?

IV. SYSTEM DESIGN AND IMPLEMENTATION

A. Architecture

The frontend uses React 19, Vite, Tailwind CSS, React Router, Recharts and Lucide React. The backend uses Python/FastAPI, SQLAlchemy, Pydantic and SQLite. REST endpoints support organisation creation, processing profiles, assessment creation, response submission, evaluation, human-review marking, AI explanations, assistant chat, regulatory records, benchmark data and report payloads. The application supports deterministic demo mode without a Gemini key and optional Gemini generation when configured. The RAG implementation loads a structured JSON knowledge base and ranks documents using direct requirement, section and title matches plus word overlap. It is therefore best described as keyword-based retrieval with generation, not a validated dense-semantic retrieval system. Embedding-based retrieval and retrieval-quality evaluation remain future work.

B. Safeguard Categories

Internal ID	Category	Examples
R6-A	Data security measures	Encryption, masking or tokens
R6-B	Access controls	Individual accounts, role restrictions, offboarding
R6-C	Logging and monitoring	Access logs, monitoring and review
R6-D	Backup and continuity	Backups, restoration testing
R6-E	Retention and disposal	Retention schedules, log retention, secure disposal
R6-F	Processor contracts	Written contracts and security clauses
R6-G	Technical and organisational measures	Policies, training, incident response

The interface includes organisation and processing profiles, an adaptive questionnaire, review and submission, dashboard, findings, human-review controls, a knowledge-base explorer, assistant, research experiment page and printable report. Outputs include Appears Addressed, Partially Addressed, Potential Gap, Unknown and Requires Human Review. These are prototype assessment labels based on reported information, not regulatory determinations.

C. Assessment Engine and Score

The deterministic engine groups responses by requirement and applies programmed rules to derive a status, priority, explanation and recommended action. The dashboard exposes a 0–100 score. The paper should call this the Security Safeguard Assessment Score and document the exact formula, applicability rules, treatment of Unknown/Not Applicable responses and category weighting from the source code. It is not an official DPDP compliance score.

D. AI Explanation and Human Review

The explanation component retrieves a requirement record and generates plain-language meaning, analysis of the reported practice, practical actions and evidence suggestions. When configured, Gemini is prompted to use retrieved context and avoid inventing legal provisions. Prompting alone cannot guarantee factual correctness or grounding, so source verification and unsupported-claim evaluation are required.

E. Security of the Prototype

The application stores organisational profiles and assessment data through SQLAlchemy models. Before public deployment or participant research, the system requires a dedicated review of authentication, access control, encryption, secret management, CORS restrictions, deletion, logging, secure hosting and threat modelling.

V. RESEARCH METHODOLOGY

A. Implemented Workflow

The current workflow collects an organisation baseline, records data types/collection method's/storage locations, gathers questionnaire responses, runs the deterministic assessment engine, displays scores and findings, provides retrieved-context explanations, permits human-review marking and generates a report payload. It relies on self-reported information and does not independently inspect infrastructure or verify evidence.

B. Current Synthetic Benchmark

The code defines 30 synthetic scenarios. The first five contain hand-authored profiles, responses and expected statuses; scenarios 6–30 are generated from sector templates. The benchmark engine compares expected labels with the rule engine. Crucially, Model B's predictions are modified using fixed index-based error rules and Model C largely reuses Model A's output. They are not actual independent zero-shot LLM and hybrid runs. The code comment referring to expert-annotated labels is not accompanied by an independent annotation protocol. Therefore, displayed metrics must not be presented as validated comparative AI results.

C. Proposed Valid Evaluation

1. Create a versioned scenario dataset with requirement, responses, expected status, rationale and evidence assumptions.
2. Have at least two qualified reviewers independently labelled scenarios, resolve disagreements and report inter-annotator agreement.

3. Freeze the test set and separate development scenarios from held-out scenarios.
4. Run Model A as the deterministic engine, Model B using an actual zero-shot LLM call, and Model C using the actual retrieval-grounded pipeline. Save model/version, prompts, retrieved passages and raw outputs.
5. Report confusion matrices, exact multi-class agreement, macro precision/recall/F1, accuracy and citation/grounding errors. Define binary gap detection separately.
6. Evaluate retrieval quality and have reviewers score whether recommendations are useful and supported.
7. Conduct a consented MSME usability pilot measuring task completion, time, abandonment, clarity and perceived usefulness.
8. Do not treat synthetic benchmark results as proof of real-world adoption, legal compliance or cost savings.

VI. CURRENT RESULTS AND VERIFICATION

The repository includes a test_verification.py script intended to test 12 capabilities, including system status, knowledgebase loading, organisation/profile creation, assessment evaluation, explanation/chat endpoints, benchmark retrieval, report generation and frontend availability. The presence of a test script does not prove that tests passed in the current environment. Run it from a clean installation and archive the actual console output before reporting a pass rate.

The README reports accuracy, precision, recall and F1 values for three model-labelled approaches. Because the implementation simulates some predictions, these values are excluded as validated empirical findings in this manuscript. The currently defensible result is the implementation of the prototype and an initial synthetic scenario generator; effectiveness remains to be evaluated.

VII. FIELD-LEVEL LIMITATIONS AND PROPOSED SOLUTIONS

Limitation / Open Problem	Why It Matters	Proposed Implementation
Synthetic labels and simulated model outputs	Metrics can overstate performance.	Actual model runs, expert labels, held-out test set and transparent metric code.
Self-reported evidence	Respondents may misunderstand or overstate controls.	Optional redacted evidence upload, evidence metadata and human validation.
Narrow legal coverage	Only Section 8(5)/Rule 6 is in scope.	Stage-wise expansion to notice, consent, rights/grievances, children and breach response after legal review.
Keyword-based retrieval	Term overlap can miss paraphrases or retrieve weak matches.	Hybrid keyword/vector retrieval, embeddings, reranking and clause-level benchmark.
No measured time/cost benefit	Affordability and time saving are currently goals, not results.	Measure task duration, review hours, API/hosting cost and comparable baseline costs.
No production security validation	The tool handles organisational security information.	Authentication, least privilege, encryption, deletion, threat modelling and penetration testing.
No live MSME usability evidence	Synthetic cases do not establish adoption.	Consented pilot with varied MSME sizes/sectors and transparent reporting.
Regulatory change management	Stale mappings can mislead users.	Versioned official sources, effective dates, change logs and expert approval.
Accessibility and digital literacy	Users may have limited legal or technical knowledge.	Mobile-first design, plain language, regional languages and assisted completion.
No remediation lifecycle	A report does not ensure gaps are fixed.	Action owner, due date, cost/effort, progress, closure evidence and reassessment history.

VIII. COST-EFFECTIVENESS, TIME SAVINGS AND MSME USABILITY

Cost-effectiveness, time saving and ease of use are central design goals, not yet measured outcomes. The guided questionnaire is intended to reduce the effort of interpreting safeguard topics; applicability logic can avoid irrelevant questions; plain-language explanations can reduce initial research time; and a consolidated report can reduce manual compilation. A deterministic demo mode also means basic assessment does not require a paid LLM API.

A future study should measure assessment completion time, reviewer validation time, direct

software/API/hosting cost, internal staff time or consultant cost for a comparable initial screening, task completion/abandonment, user-rated clarity and willingness to pay. Savings should be calculated only from observed data and a disclosed baseline. The tool should be positioned as low-cost preliminary decision support that may help an MSME prepare for professional review, not as a replacement for legal advice or independent security audits.

Affordability options include local deterministic operation, cached regulatory content, low-cost model options, usage limits, exportable checklists, a free starter tier and partnerships with MSME associations or academic/legal support groups. The product should minimise collection of business details and should not require raw personal records for a basic assessment.

IX. FUTURE SCOPE

- Expand the framework beyond Rule 6 to relevant DPDP obligations, including notice and consent, data-principal rights, grievance handling, children's data and breach response, while representing applicability and commencement dates.
- Develop validated hybrid RAG with versioned official sources, semantic embeddings, keyword/vector retrieval, reranking and clause-level citations.
- Add evidence-assisted assessment using redacted policy extracts or configuration summaries, with human confirmation of evidence sufficiency.
- Add a remediation planner with action owner, due date, effort/cost estimate, status tracking and reassessment history.
- Run independent comparative experiments with expert-labelled scenarios and actual rule-based, zero-shot LLM and hybrid outputs.
- Conduct an MSME pilot measuring usability, time savings, affordability, trust and perceived usefulness.
- Add regional-language explanations, mobile-first flows and assisted completion for users with limited technical/legal literacy.
- Build a regulatory update pipeline with official-source monitoring, version control, expert approval and impact analysis.
- Investigate a sustainable low-cost model: free self-assessment, optional paid assisted review, affordable subscription for repeated assessments or institutional partnerships.

X. ETHICAL, LEGAL AND SECURITY CONSIDERATIONS

Collect only information necessary for the assessment, avoid requesting actual personal records when descriptions suffice, and provide clear retention and deletion choices. Human-participant studies should use informed consent, anonymisation/pseudonymisation, secure storage, restricted access and any university ethics approval required. AI outputs should be labelled, cite the relevant source, acknowledge uncertainty and permit human correction. A high score does not prove compliance; a low score is not a legal determination.

XI. CONCLUSION

DPDP Sentinel implements a software workflow for structured, AI-assisted personal-data security self-assessment focused on Section 8(5) of the DPDP Act and Rule 6 safeguard categories. Its practical contribution is the integration of a questionnaire, deterministic assessment, structured regulatory records, explanatory assistance, human-review marking and report output. Code inspection also identifies a methodological issue: the current comparative benchmark simulates some model outputs and cannot substantiate claims of superior AI performance. The next phase should priorities independent expert annotation, actual model execution, retrieval evaluation, security review and an MSME usability pilot. These steps will determine whether the design is accurate, usable, affordable and time-saving in practice.

REFERENCES

[1]. Government of India, The Digital Personal Data Protection Act, 2023, Act No. 22 of 2023, India Code.

[2]. Ministry of Electronics and Information Technology, Government of India, Digital Personal Data Protection Rules, 2025, Gazette of India, G.S.R. 846(E).

[3]. Ministry of Electronics and Information Technology, Government of India, Explanatory Note to the Digital Personal Data Protection Rules, 2025.
[4]. J. Kingston, "Using Artificial Intelligence to Support Compliance with the General Data Protection Regulation," Artificial Intelligence and Law, vol. 25, no. 4, pp. 429–443, 2017.
[5]. F. Lorè, P. Basile, A. Appice, M. de Gemmis, D. Malerba, and G. Semeraro, "An AI framework to support decisions on GDPR compliance," Journal of Intelligent Information Systems, vol. 61, pp. 541–568, 2023.
[6]. S. Tarafder et al., "Informational Privacy in the Age of Artificial Intelligence: A Critical Analysis of India's DPDP Act, 2023," Panjab University Law Review, vol. 65, no. 1, pp. 112–128, 2024.
[7]. R. Sharma and S. Tarafder, "Compliance Automation and Statutory Reasoning for Emerging Data Protection Regimes," Journal of Information Economics and Law, vol. 12, no. 2, pp. 45–68, 2025.
[8]. A. Verma, "Evaluating Hybrid Retrieval-Augmented Generation Models in Legal Document Processing," IEEE Transactions on Emerging Topics in Computing, 2026.
[9]. K. Patel et al., "Assessing Security Safeguards for Small and Medium Enterprises under Modern Privacy Mandates," Applied Artificial Intelligence, vol. 40, no. 1, Art. no. e2626097, 2026.
[10]. P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," in Advances in Neural Information Processing Systems, vol. 33, pp. 9459–9474, 2020.
[11]. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust you? Explaining the Predictions of Any Classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.
[12]. I. Chalkidis, A. Jana, E. Dirani, M. Zampieri, I. Androutsopoulos, and N. Aletras, "LexGLUE: A Benchmark Dataset for Legal Language Understanding in English," in Proc. 60th Annu. Meeting Assoc. Comput. Linguistics, 2022, pp. 4310–4330.
[13]. S. Zimmeck, P. Story, A. Smullen, A. Ravichander, Z. Wang, J. Reidenberg, N. Russell, and N. Sadeh, "MAPS: Scaling Privacy Compliance Analysis to a Million Apps," Proceedings on Privacy Enhancing Technologies, vol. 2019, no. 3, pp. 66–86, 2019.
[14]. G. Greenleaf, "India's Data Protection Act, 2023: A New Model for the Global South?" Privacy Laws & Business International Report, vol. 185, pp. 1–6, 2023.
[15]. A. K. Sharma and M. Kaur, "Cybersecurity Challenges and Framework Adoption in Indian MSMEs: An Empirical Study," International Journal of Information Management Data Insights, vol. 4, no. 2, Art. no. 100250, 2024.
[16]. S. Esayas, "The role of anonymisation and pseudonymisation under the EU data protection rules," International Journal of Law and Information Technology, vol. 23, no. 3, pp. 252–275, 2015.
[17]. P. Hacker, "A Legal Framework for AI Training Data," Law, Innovation and Technology, vol. 13, no. 2, pp. 257–301, 2021.
[18]. N. S. Nappinai, "Data Privacy in India: Evolution and the DPDP Act 2023," Indian Journal of Law and Technology, vol. 19, no. 1, pp. 45–72, 2024.
[19]. R. B. Basyal, S. Nepal, and R. K. Ko, "Cloud Security and Privacy for Small and Medium Enterprises," IEEE Cloud Computing, vol. 8, no. 3, pp. 24–33, 2021.
[20]. H. Nayak, S. Deshmukh, and V. Singh, "Domain-Specific Language Models for Indian Legal Context: Challenges and Opportunities," in Proc. ACM India Joint Int. Conf. Data Science and Management of Data (CoDS-COMAD), 2025, pp. 142–150.
[21]. S. Eshelman, K. B. Anwar, and D. L. Geffen, "RAGAS: Automated Evaluation of Retrieval Augmented Generation," arXiv preprint arXiv:2309.15217, 2023.
[22]. C. M. B. E. R. de Silva and C. S. de Silva, "Towards Automated GDPR Compliance Checking for SMEs," IEEE Access, vol. 10, pp. 110452–110467, 2022.
[23]. M. Palmirani, P. Casanovas, and U. Pagallo, "Legal Ontologies for Evaluating the Compliance of Data Processing with the DPDP Act 2023," Artificial Intelligence and Law, vol. 33, no. 2, pp. 195–218, 2025.

[24]. A. Agrawal and P. Tiwari, "Evaluating the Cost of Privacy Compliance in Developing Economies: Evidence from Indian Startups," Journal of Global Information Management, vol. 34, no. 4, pp. 1–22, 2026.

[25]. L. Floridi and M. Taddeo, "What is Data Ethics?" Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 374, no. 2083, Art. no. 20160360, 2016.