

Legal Compliance by Design: Developing a Software-Based Framework for Privacy and Data Protection in AI Applications

Dr. Jesmin Banu Khan¹, Md Rishad Khan²

¹ Associate Professor, School of Law, G H Raison University, Saikheda

² Assistant Professor, BCA Dept. G H Raison Skill Tech University, Nagpur

Abstract

Artificial Intelligence (AI) has become deeply integrated into contemporary digital life. AI-powered applications are increasingly used for recruitment, education, healthcare, banking, e-commerce, financial assessment, social networking, public administration, cybersecurity, and decision support. The effectiveness of these applications depends largely upon the availability and processing of digital data. While data-driven technologies create substantial economic and social benefits, they also create difficult questions concerning privacy, data protection, transparency, cybersecurity, accountability, discrimination, profiling, and individual autonomy. These concerns become more complex when personal data is not merely stored but continuously analysed, inferred, transformed, replicated, and incorporated into machine-learning systems.

Traditional approaches to privacy compliance frequently treat law and software development as separate activities. Legal teams interpret legislation and prepare policies, notices, contracts, and compliance documentation, while software developers translate functional requirements into databases, algorithms, interfaces, APIs, and applications. This separation creates a significant implementation gap. A privacy policy may promise responsible processing, but the software may still collect excessive information, retain it indefinitely, transmit it to third-party AI services, or use it for purposes that were not anticipated at the time of collection. The central argument of this paper is therefore that privacy and legal compliance should be incorporated directly into software architecture rather than being treated as a post-development audit activity.

This paper proposes a **Legal Compliance by Design (LCbD)** framework for AI applications. The framework extends the established concept of Privacy by Design by translating legal obligations into software requirements, technical controls, automated tests, compliance rules, audit evidence, and continuous monitoring mechanisms. The study adopts an interdisciplinary research design combining doctrinal legal analysis, privacy engineering, software requirements engineering, AI governance, cybersecurity principles, and design science. India's Digital Personal Data Protection Act, 2023 and Digital Personal Data Protection Rules, 2025 form the principal legal context. Comparative reference is made to the European Union's General Data Protection Regulation (GDPR), particularly its data-protection-by-design requirements, and to the NIST Privacy Framework and AI Risk Management Framework.

The proposed architecture contains ten principal components: a Legal Requirement Mapping Engine; Data Inventory and Classification Module; Consent and Notice Manager; Purpose and Processing Controller; Data Retention and Deletion Engine; Data Principal Rights Manager; AI Dataset Governance Module; Privacy and Security Risk Engine; Compliance Audit Dashboard; and Continuous Legal Monitoring System. The paper further proposes a legal-technical traceability matrix connecting a legal obligation to a software requirement, technical control, verification test, and audit evidence.

The findings indicate that compliance automation can substantially improve traceability, consistency, accountability, and early identification of privacy risks, but cannot completely automate legal interpretation. Human oversight remains essential, particularly where statutory provisions are context-dependent or where AI processing creates novel risks. The paper concludes that Legal Compliance by Design can provide a practical bridge between legal principles and software engineering and can contribute to a more privacy-preserving, accountable, secure, and trustworthy AI ecosystem.

Keywords: Artificial Intelligence; Privacy; Data Protection; Legal Compliance by Design; Privacy by Design; Software Engineering; Digital Personal Data Protection Act; Cybersecurity; AI Governance; Privacy Engineering; Data Governance; India.

I. Introduction:

Background and Problem Context

The rapid development of digital technologies has changed the nature of privacy. Privacy was once largely understood as the protection of physical spaces, private communications, confidential records, or personal information held by institutions. In contemporary digital environments, however, privacy is increasingly concerned with what can be **inferred, predicted, combined, and generated** from data.

An individual may provide an application with a name, telephone number, photograph, location, purchase history, or browsing behaviour. Artificial Intelligence can subsequently analyse those seemingly ordinary pieces of information to generate additional conclusions about the individual. The technological capacity to infer characteristics from data creates a significant legal challenge because privacy risks may emerge not only from what an individual directly discloses but also from what a system derives from that disclosure.

AI has intensified this challenge.

Modern applications can use machine learning to classify users, recommend products, assess creditworthiness, identify faces, detect fraud, analyse language, evaluate employment candidates, personalise education, or generate content. Generative AI adds another dimension because users may voluntarily or involuntarily submit personal information into conversational interfaces, document-analysis tools, coding assistants, or automated decision-support systems.

The fundamental problem is therefore not simply that organisations possess personal data. It is that **software systems increasingly determine what happens to that data**.

India's Digital Personal Data Protection Act, 2023 (DPDP Act) represents a significant legislative development in this environment. The Act establishes a framework for processing digital personal data while recognising both individual interests in protecting personal data and the need for lawful processing. The Digital Personal Data Protection Rules, 2025 were subsequently notified on November 14, 2025. The Rules contain a phased commencement structure, meaning that different provisions become operational at different times rather than all provisions taking effect simultaneously.

This evolving legal environment creates an important question for computer applications and software engineering: **How can legal privacy requirements be translated into actual software behaviour throughout the lifecycle of an AI application?**

A conventional compliance model might operate as follows:

Law → Policy → Legal Review → Audit

The proposed model instead operates as:

Law → Legal Requirement → Software Requirement → Technical Control → Testing → Monitoring → Evidence

The distinction is substantial.

A privacy policy may state that personal data will be protected. Legal Compliance by Design asks where, exactly, in the software architecture that protection is implemented.

For example, if an organisation states that personal data will be retained only as long as necessary, a compliance-oriented software system should have an actual retention mechanism. If an individual has a right to request action concerning their data, the application should contain an operational process through which that request can be received, verified, processed, and recorded.

Privacy therefore becomes an **engineering property**, not merely a documentation exercise.

Dimension of the Problem

The technical language surrounding data can sometimes make privacy concerns appear abstract.

A developer may see:

```
user_id = 45821
```

location = 12.9716, 77.5946

purchase_history = [...]

model_score = 0.87

But these are not merely database fields.

Behind each record is a human being.

A location record may reveal where someone lives or works. A purchase history may reveal intimate preferences.

A photograph may reveal identity. A voice recording may be associated with a person for years. A behavioural profile may influence employment, financial services, education, insurance, or access to digital platforms.

Consequently, privacy engineering should retain a human dimension. Data protection is ultimately about protecting **people, choices, dignity, autonomy, and trust.**

Research Problem

The central research problem addressed by this paper is the gap between:

1. legal privacy obligations;
2. organisational privacy policies;
3. software requirements;
4. AI development practices; and
5. operational technical controls.

Existing privacy programmes may be documentation-heavy, while software development may remain focused primarily on functionality, performance, usability, and security. AI development introduces additional complexity through training datasets, model pipelines, data replication, inference, third-party APIs, and continuous model improvement.

There is consequently a need for a framework that can connect the legal and digital dimensions.

Background of the Study

Growth of Artificial Intelligence Applications

AI has become increasingly embedded in everyday digital services. Search engines, online marketplaces, banking applications, digital assistants, social networks, healthcare systems and educational platforms use AI to analyse information and generate predictions.

AI applications can broadly involve several categories of data:

- identity data;
- contact information;
- financial information;
- behavioural information;
- location information;
- biometric information;
- images and audio;
- educational records;
- employment information;
- online activity;
- transaction records; and
- inferred or predicted information.

The last category is especially important.

AI does not merely process information supplied directly by an individual. It can derive new information from existing information. For example, an application may infer a person's interests, financial risk, behavioural characteristics or preferences.

This creates a legal difficulty because the privacy impact of AI cannot always be understood simply by examining the original database. The processing pipeline itself may create additional privacy risks.

Statement of the Problem

The traditional software development process often treats legal compliance as an external activity.

A simplified lifecycle may look like this:

Requirement Analysis → Design → Development → Testing → Deployment → Maintenance

Privacy compliance is frequently added as an additional activity:

Software Development + Legal Review

This model is inadequate for AI applications.

A more appropriate model should integrate legal requirements throughout the lifecycle:

Legal Analysis → Data Mapping → Technical Requirements → Architecture → Development → Privacy Testing → Deployment → Continuous Compliance Monitoring

The central problem addressed by this research is therefore the absence of an easily implementable framework that translates legal requirements into technical controls within AI software development.

Legal professionals understand legislation but may not understand the technical architecture of machine-learning systems. Software developers understand databases, APIs, algorithms and cloud infrastructure but may not understand the legal meaning of purpose limitation, lawful processing, notice, consent, retention or individual rights.

The result is a communication gap.

LCbD attempts to bridge that gap.

Objectives

The primary objective of this study is to develop a conceptual **Legal Compliance by Design framework for AI applications**.

The specific objectives are:

- i. To analyse privacy and data-protection risks arising from AI applications.
- ii. To examine the relevance of India's DPDP framework to software development.
- iii. To critically analyse existing Privacy by Design and AI-risk-management approaches.
- iv. To develop a methodology for translating legal requirements into software requirements.
- v. To propose a software architecture supporting continuous compliance.
- vi. To develop algorithms for legal-rule mapping, data-flow analysis, risk scoring, retention monitoring, and compliance traceability.
- vii. To examine the role of cybersecurity and privacy-enhancing technologies.
- viii. To identify the limitations of automated legal compliance.
- ix. To provide recommendations for organisations developing or deploying AI applications.

1.5 Research Questions

The study addresses six principal questions:

RQ1: What privacy and data-protection risks are distinctive to AI applications?

RQ2: What legal requirements should be reflected in the design of AI applications?

RQ3: How can legal requirements be converted into functional and non-functional software requirements?

RQ4: What technical architecture can support continuous privacy compliance?

RQ5: Can compliance monitoring be partially automated without eliminating human legal judgment?

RQ6: How can a Legal Compliance by Design framework improve accountability and traceability?

II. Literature Review:

Privacy as a Legal and Technological Concept

Privacy scholarship has traditionally developed across law, philosophy, information science, and technology. Warren and Brandeis's classic conception of privacy as a right to be let alone established an influential foundation for privacy theory (Warren and Brandeis 1890). Later scholarship expanded privacy beyond physical intrusion to include informational control, contextual integrity, autonomy, identity, and the appropriate flow of information (Westin 1967; Nissenbaum 2010).

The technological transformation of privacy demonstrates why purely legal or purely technical approaches are insufficient.

Information can now be collected passively, processed automatically, combined across systems, and analysed at a scale that was previously impossible. Consequently, privacy protection increasingly requires technical mechanisms capable of implementing legal principles.

Solove's work on privacy emphasises the difficulty of reducing privacy to a single concept, instead identifying a broad range of problematic information practices including collection, processing, dissemination, and invasion (Solove 2006). This is especially relevant to AI because AI applications may engage in all four simultaneously.

Privacy by Design

The concept of **Privacy by Design (PbD)** is one of the most influential approaches for integrating privacy into technological systems.

Cavoukian's formulation emphasised proactive rather than reactive privacy protection and advocated embedding privacy into system design rather than attempting to repair privacy problems after deployment (Cavoukian 2011).

The European Union's GDPR subsequently gave Privacy by Design a strong legal dimension. Article 25 requires controllers to implement appropriate technical and organisational measures designed to implement data-protection principles and integrate safeguards into processing. It expressly refers to measures such as pseudonymisation and data minimisation.

Article 25 is particularly significant for the present research because it establishes a direct relationship between **law and system design**.

The implication is not simply that organisations should have privacy policies. Rather, appropriate technical and organisational measures should be incorporated when determining the means of processing and during processing itself.

Requirements Engineering and Privacy

Software requirements engineering traditionally focuses on identifying what a system should do and the constraints under which it should operate.

Privacy introduces a special category of requirements.

A requirement such as:

"The system shall allow users to log in."

is relatively straightforward to test.

A privacy requirement such as:

"The system shall process personal data only for authorised purposes."

is more complex.

It requires an understanding of:

- what personal data exists;
- what purpose was specified;
- what processing is occurring;
- who authorised the processing;
- whether the purpose has changed;
- what legal requirements apply; and
- whether the technical system enforces the restriction.

Recent research directly addresses this problem. Kosenkov et al. (2026) examined the alignment of GDPR requirements with software engineering specifications and argued that legal knowledge should be captured at

multiple abstraction levels throughout the engineering lifecycle. Their research emphasises traceability and the relationship between legal requirements and system specifications.

This work is highly relevant but also identifies an opportunity for further research.

Much of the existing literature is strongly influenced by the GDPR. There remains scope for an India-focused model that connects the DPDP framework with **AI-specific data pipelines, compliance automation, and software architecture.**

Privacy Engineering

Privacy engineering attempts to operationalise privacy principles through technical methods.

Common mechanisms include:

- data minimisation;
- pseudonymisation;
- anonymisation;
- encryption;
- access control;
- privacy-preserving analytics;
- differential privacy;
- federated learning;
- secure computation; and
- data lifecycle management.

The NIST Privacy Framework provides a structured approach for improving privacy through enterprise risk management. NIST describes the framework as a tool rather than a binding law and positions it as a means of managing privacy risks at organisational level.

The strength of privacy engineering is its technical orientation. Its limitation is that technical controls alone cannot determine the meaning of legal obligations.

Encryption, for example, may be technically excellent but does not answer whether an organisation was legally entitled to collect the data in the first place.

AI Risk Management

AI governance has developed alongside privacy engineering.

The NIST AI Risk Management Framework (AI RMF) is designed to help organisations manage AI risks across the design, development, deployment, and evaluation lifecycle. It is voluntary, rights-preserving, non-sector-specific, and use-case agnostic.

The framework uses four major functions:

- **Govern**
- **Map**
- **Measure**
- **Manage**

The accompanying Playbook provides suggested actions for operationalising those functions.

The NIST Generative AI Profile extends this approach to risks associated specifically with generative AI.

The AI RMF is valuable because it demonstrates how AI governance can be structured as a lifecycle process rather than a one-time review.

However, AI risk management and legal compliance are not identical.

AI risk management asks:

What could go wrong?

Legal compliance asks:

What does the law require?

Legal Compliance by Design attempts to combine the two.

Data Protection and AI

AI creates privacy problems that traditional database systems may not encounter.

Training-data exposure

A model may be trained on large datasets containing personal information.

Data duplication

Training datasets may be copied into multiple storage environments.

Model memorisation

Some models may retain or reproduce information from training data.

Inference

A model may derive characteristics that an individual never explicitly provided.

Third-party processing

An organisation may send personal information to an external AI service.

Continuous learning

A model may be updated periodically using newly collected data, creating an ongoing compliance problem.

These issues demonstrate that privacy must follow the data throughout the AI lifecycle.

Research Gap

The literature reveals four important gaps.

First, privacy-by-design research has established the need to integrate privacy into software development, but implementation remains difficult because legal requirements are expressed at a different level of abstraction from technical specifications.

Second, AI risk frameworks address trustworthy AI but are generally broader than data-protection law.

Third, data-protection law establishes rights and obligations but does not provide developers with a complete software architecture for implementing those obligations.

Fourth, there is a need for an India-focused model that translates the DPDP framework into **software requirements and AI lifecycle controls**.

The present study addresses these gaps through the proposed LCbD architecture.

III. Methodology: Research Design and Algorithms**Research Methodology**

This study follows a **doctrinal, interdisciplinary and design-oriented research methodology**.

Doctrinal Legal Research

The legal component examines:

- Digital Personal Data Protection Act, 2023;
- Digital Personal Data Protection Rules, 2025;
- constitutional privacy principles;
- information technology law;
- cybersecurity requirements;
- relevant international privacy frameworks;
- principles associated with privacy by design; and
- emerging AI governance principles.

The DPDP Act applies to processing of digital personal data within India under specified circumstances and can also apply to certain processing outside India connected with offering goods or services to Data Principals in India.

Technical Research

The technical component examines:

- software development lifecycle;
- database architecture;
- machine-learning pipelines;
- data classification;
- access control;
- encryption;
- anonymisation and pseudonymisation;
- API security;
- logging;
- audit trails;

- consent management;
- data deletion;
- model governance; and
- privacy-preserving computation.

Design Science Approach

The proposed framework is treated as a design artefact.

The research therefore moves from:

Legal Problem → Requirement → Technical Control → Software Module → Compliance Evidence

This approach is particularly suitable because recent software-engineering research has demonstrated the value of connecting legal requirements with software specifications and traceability. A 2026 study on privacy-by-design found that legal knowledge needs to be represented across different levels of software abstraction to support effective compliance.

Legal Perspective

Constitutional Dimension of Privacy in India

Privacy is not merely a contractual interest. It has constitutional significance in India.

In *Justice K.S. Puttaswamy (Retd.) v. Union of India*, the Supreme Court recognised privacy as a constitutionally protected right under Article 21 and related constitutional guarantees.

The judgment changed the legal understanding of privacy in India by recognising privacy as essential to dignity, autonomy and liberty.

This is particularly relevant to AI because AI systems can influence individuals without their direct awareness.

An individual may know that a bank collects transaction information. The individual may not know that a machine-learning model is using thousands of historical variables to infer creditworthiness.

Similarly, a social media user may know that the platform records activity. The user may not understand that the platform can use AI to infer interests, personality characteristics or behavioural patterns.

The constitutional dimension therefore reinforces the need for transparency and accountability in AI applications.

Digital Personal Data Protection Act, 2023

The DPDP Act represents a significant development in India's data-protection framework.

The Act uses the concepts of **Data Principal**, **Data Fiduciary** and **Data Processor**.

From a software perspective, these concepts can be mapped onto system architecture.

For example:

Data Principal → Application User

Data Fiduciary → Organisation determining purpose and means

Data Processor → Cloud provider / analytics provider / AI service provider

This mapping is valuable because software systems can be designed to record the relationship between these actors.

The Act includes provisions concerning processing, consent, notice, obligations of Data Fiduciaries, rights of Data Principals, security safeguards, breach-related obligations and penalties.

A software framework should therefore not merely store the text of the law. It should convert these requirements into operational controls.

Digital Personal Data Protection Rules, 2025

The notification of the DPDP Rules, 2025 represents an important stage in the operationalisation of India's privacy regime.

The Government notified the Rules on 14 November 2025 after a public consultation process. The Government stated that the Rules give effect to the DPDP Act and establish operational mechanisms for responsible processing and protection of digital personal data.

MeitY's official materials list the final Rules, the enforcement timeline and material concerning the Data Protection Board.

This development has direct implications for software architecture.

A company cannot reasonably treat privacy compliance as a static policy document when its applications continuously collect, process, transfer and delete data.

Instead, the application should maintain a dynamic compliance state.

For example:

User Data Collected → Purpose Recorded → Legal Basis/Consent Recorded → Processing Permitted → Retention Timer Activated → Rights Requests Supported → Deletion/Erasure Executed → Audit Evidence Created

This is the core philosophy of Legal Compliance by Design.

Privacy by Design and Its Evolution into Legal Compliance by Design

Privacy by Design traditionally means incorporating privacy considerations into the design and architecture of systems rather than adding them after development.

The approach is closely associated with principles such as:

1. proactive rather than reactive protection;
2. privacy as a default setting;
3. privacy embedded into design;
4. full functionality;
5. end-to-end security;
6. visibility and transparency; and
7. respect for user privacy.

The concept is highly relevant to software engineering.

However, this paper argues that **Privacy by Design should be expanded into Legal Compliance by Design**.

The distinction is important.

Privacy by Design asks:

How can we design software that protects privacy?

Legal Compliance by Design asks:

How can we design software that operationalises legally applicable privacy obligations and produces evidence that those obligations are being satisfied?

The second question is broader.

It connects law with:

- requirements engineering;
- software architecture;
- databases;
- APIs;
- security controls;
- testing;
- monitoring;
- auditing; and
- governance.

Recent research confirms the growing importance of translating privacy law into software specifications. A 2026 Information and Software Technology study specifically examined the alignment of GDPR requirements with requirements engineering and system specifications.

The proposed framework extends this idea into the Indian AI context.

Research Design

The study uses a **multidisciplinary design-science methodology** combining:

1. doctrinal legal research;
2. systematic conceptual literature analysis;

3. requirements engineering;
4. privacy engineering;
5. AI lifecycle analysis;
6. cybersecurity analysis; and
7. software architecture design.

The research process consists of five stages:

Stage 1: Identify legal obligations.

Stage 2: Translate obligations into software requirements.

Stage 3: Map requirements to technical controls.

Stage 4: Develop algorithms for compliance monitoring.

Stage 5: Evaluate the conceptual framework through hypothetical AI application scenarios.

The study is primarily a **design and conceptual evaluation**, rather than a claim of empirical deployment across real organisations. Accordingly, the Results section reports findings from the framework analysis and scenario-based evaluation rather than invented operational performance statistics.

Legal Corpus

The primary legal corpus comprises:

- Digital Personal Data Protection Act, 2023;
- Digital Personal Data Protection Rules, 2025;
- relevant Indian constitutional privacy jurisprudence;
- relevant information-technology law; and
- selected comparative frameworks.

The DPDP Act itself states that its purpose is to provide for processing of digital personal data in a manner recognising individual rights and the need for lawful processing.

The 2025 Rules are particularly important because they include provisions relating to techno-legal measures and establish a phased commencement structure.

Conceptual Architecture

The proposed LCbD system consists of ten modules:

1. **Legal Requirement Mapping Engine**
2. **Data Inventory and Classification Module**
3. **Consent and Notice Manager**
4. **Purpose and Processing Controller**
5. **Retention and Deletion Engine**
6. **Data Principal Rights Manager**
7. **AI Dataset Governance Module**
8. **Privacy and Security Risk Engine**
9. **Compliance Audit Dashboard**
10. **Continuous Legal Monitoring System**

The architecture follows the following logic:

Legal Source

↓

Legal Requirement

↓

Software Requirement

↓

Technical Control

↓

Compliance Test

↓

Evidence



Continuous Monitoring

Legal Requirement Mapping Algorithm

The first algorithm converts legal requirements into structured software requirements.

Algorithm 1: Legal-to-Technical Requirement Mapping

INPUT:

Legal_Text
Processing_Context
Application_Architecture

STEP 1:

Extract legal obligations from Legal_Text.

STEP 2:

Classify each obligation as:

- Notice
- Consent
- Purpose
- Security
- Retention
- Rights
- Accountability
- Governance

STEP 3:

Identify affected processing activities.

STEP 4:

Convert each obligation into a software requirement.

STEP 5:

Map each software requirement to one or more technical controls.

STEP 6:

Assign verification tests.

STEP 7:

Create traceability record.

OUTPUT:

Legal Requirement → Software Requirement
→ Technical Control → Test → Evidence

For example:

Legal principle: Protect personal data through appropriate safeguards.

Software requirement: Personal-data repositories must enforce approved security controls.

Technical control: Encryption + access control + monitoring.

Test: Verify configuration and attempt unauthorised access.

Evidence: Test result and audit record.

Data Classification Algorithm

The system should identify personal data across databases and AI datasets.

Algorithm 2: Data Classification

FOR each data_element IN system:

identify data source

determine whether data relates to an identifiable individual

IF personal:

assign classification = PERSONAL

IF sensitive/high-risk according to applicable policy:

assign classification = HIGH_RISK

record:

purpose

owner

processor

retention

access_level

AI_usage

END FOR

This algorithm creates a data inventory.

Purpose Compatibility Algorithm

AI systems often reuse data.

The proposed purpose controller therefore evaluates new processing.

INPUT:

Original_Purpose

Proposed_Purpose

Data_Category

Legal_Rules

Consent_Status

IF Proposed_Purpose == Original_Purpose:

permit processing subject to controls

ELSE:

evaluate legal conditions

IF additional authorisation/notice required:

flag for review

IF processing prohibited:

BLOCK

ELSE:

permit subject to documented approval

The system should never treat algorithmic approval as a substitute for legal judgment.

Privacy Risk Scoring Algorithm

A practical compliance system needs a risk score.

A conceptual model is:

Risk = Likelihood × Impact × Exposure

Where:

- **Likelihood** represents probability of misuse or breach;
- **Impact** represents potential harm;
- **Exposure** represents the scale and sensitivity of processing.

A weighted model can be expressed as:

$$R = 0.35L + 0.40I + 0.25E$$

where each factor is normalised to a 0–10 scale.

The weighting is illustrative rather than a legally mandated formula.

The system could classify:

0–2.9 = Low

3.0–5.9 = Moderate

6.0–7.9 = High

8.0–10 = Critical

Critical cases should require human review.

Data Retention Algorithm

FOR each personal_data_record:

 retrieve collection_date

 retrieve approved_retention_period

 calculate expiry_date

 IF current_date >= expiry_date:

 flag record

 IF legal_hold == FALSE:

 initiate deletion workflow

 ELSE:

 preserve and document legal hold

END FOR

For AI applications, the algorithm should also inspect:

- source database;
- training dataset;
- feature store;
- model-development environment;
- backup systems;
- embeddings;
- cached files.

This is necessary because deleting information from one database does not necessarily remove it from the entire AI ecosystem.

Dataset Governance Algorithm

INPUT:

 Dataset

CHECK:

 Source documented?

 Purpose documented?

 Personal data identified?

 Legal status recorded?

Retention defined?

Security classification assigned?

Approved for AI training?

IF any critical requirement == FALSE:

TRAINING_STATUS = BLOCKED

ELSE:

TRAINING_STATUS = APPROVED

Create audit record.

This makes dataset governance a formal stage in the machine-learning pipeline.

IV. Results: Objective Presentation of Findings

Finding 1: Legal Requirements Can Be Represented as Software Controls

The analysis demonstrates that many privacy requirements can be translated into technical controls.

Legal/Policy Requirement	Software Control	Verification
Transparency	Notice Manager	Notice-version test
Consent management	Consent database	Consent audit
Purpose control	Processing controller	Purpose validation
Data security	Encryption/RBAC	Security test
Retention	Automated deletion	Retention test
Rights management	Privacy portal	Request workflow test
Accountability	Audit logging	Evidence inspection
AI governance	Dataset passport	Dataset approval test

The important finding is that legal obligations and technical controls are not identical, but they can be connected through a traceability layer.

Finding 2: Compliance Requires Data Lineage

The framework demonstrates that AI compliance cannot be limited to the primary production database.

A complete data path may be:

Individual

↓

Application

↓

Production Database

↓

Data Warehouse

↓

Training Dataset

↓

Model

↓

Prediction

↓

Third-Party API

Every stage creates a potential privacy exposure.

Consequently, an effective LCbD system must maintain data lineage.

Finding 3: AI Creates Additional Compliance Points

Traditional software may require compliance at collection and storage.

AI introduces additional stages:

- dataset creation;
- preprocessing;
- feature engineering;
- training;
- validation;
- model deployment;
- inference;
- model updates;
- monitoring;
- retirement.

This means the compliance lifecycle must mirror the AI lifecycle.

Finding 4: Automated Compliance Is Feasible but Limited

The analysis indicates that some compliance activities are highly suitable for automation.

These include:

- detecting missing encryption;
- identifying unclassified data;
- checking retention dates;
- recording consent;
- identifying unauthorised access;
- testing API permissions;
- maintaining audit logs;
- checking dataset metadata;
- flagging missing documentation.

Other activities are less suitable for complete automation.

These include:

- interpreting ambiguous legal provisions;
- determining proportionality;
- resolving conflicts between legal obligations;
- assessing complex contextual risks;
- interpreting new judicial decisions.

The appropriate model is therefore **human-supervised automation**.

Finding 5: Traceability Improves Accountability

The proposed legal-technical traceability matrix provides an answer to two fundamental audit questions:

What legal requirement are we trying to satisfy?

and

How does our software satisfy it?

For example:

ID	Legal Requirement	Software Component	Control	Evidence
LC-01	Privacy notice	Notice Manager	Version control	Notice log
LC-02	Consent	Consent Manager	Timestamped consent	Consent record
LC-03	Security	Security Layer	Encryption/RBAC	Test report
LC-04	Retention	Retention Engine	Auto-delete	Deletion log
LC-05	Rights	Rights Portal	Request workflow	Request record
LC-06	AI governance	Dataset Module	Dataset approval	Dataset passport

This traceability is consistent with the direction of contemporary privacy-engineering research, which emphasises connecting legal concepts with requirements and system specifications.

Finding 6: Privacy and Cybersecurity Are Interdependent

The framework also demonstrates that privacy and cybersecurity cannot be completely separated.

A weak access-control system can expose personal information.

An insecure API can transmit personal information to unauthorised parties.

Poor credential management can compromise an AI model.

Inadequate logging can make it impossible to determine what happened after an incident.

Therefore, privacy controls should be integrated into the broader security architecture.

Finding 7: Existing Frameworks Are Complementary

The analysis suggests that no single framework completely addresses the problem.

The GDPR provides strong legal principles concerning privacy by design.

The NIST Privacy Framework provides enterprise privacy-risk-management guidance.

The NIST AI RMF provides AI-risk-management functions and trustworthy-AI considerations.

The DPDP framework provides the principal Indian statutory context.

The proposed LCbD framework attempts to connect these domains rather than replace them.

V. Discussion: Interpretation and Implications

From Privacy Policy to Privacy Architecture

One of the most important implications of this research is the need to rethink the role of privacy policies.

A privacy policy remains important, but it should not be the endpoint of compliance.

Consider an application stating:

“We retain personal data only for as long as necessary.”

The statement is meaningful only if the underlying system has a mechanism for determining:

- what data was collected;
- why it was collected;
- when it was collected;
- how long it should be retained;
- whether a legal hold applies;
- whether it has expired; and
- whether deletion was completed.

The software therefore becomes the operational layer of the legal policy.

Law as a Software Requirement

Traditional software development often begins with business requirements.

LCbD adds another category:

Legal Requirements.

The requirements hierarchy becomes:

Business Requirements

What should the system achieve?

Functional Requirements

What should the system do?

Non-functional Requirements

How should the system perform?

Security Requirements

How should the system be protected?

Legal Requirements

What must the system do or avoid doing because of applicable law?

Ethical Requirements

What should the system do to protect human interests beyond minimum legal compliance?

This expanded requirements model is especially important for AI.

Legal Requirements Are Not Always Binary

A major challenge is that law is contextual.

A software rule might say:

IF consent = TRUE

THEN processing = ALLOWED

But legal compliance is rarely that simple.

Consent may be:

- valid or invalid;
- informed or uninformed;
- withdrawn or active;
- purpose-specific;
- insufficient for a particular processing activity.

Similarly, security cannot be reduced to:

encryption = TRUE

because security depends on the broader threat model.

LCbD must therefore be understood as a **decision-support and control architecture**, not a machine that mechanically determines legal truth.

Human-in-the-Loop Legal Compliance

The framework proposes three levels of automation.

Level 1: Automatic Controls

Suitable for:

- encryption checks;
- access-control checks;
- retention alerts;
- logging;
- consent recording.

Level 2: Automated Risk Detection

Suitable for:

- identifying unusual processing;
- detecting data flows;
- flagging high-risk datasets;
- identifying missing compliance evidence.

Level 3: Human Legal Judgment

Required for:

- ambiguous interpretation;
- complex statutory questions;
- proportionality;
- novel AI practices;
- conflicting legal obligations.

This creates the following model:

Automation → Risk Flag → Human Review → Decision → Technical Enforcement

Implications for Software Developers

Software developers should increasingly understand privacy as part of professional responsibility.

A developer does not need to become a lawyer.

However, developers should understand concepts such as:

- personal data;
- data minimisation;
- purpose limitation;
- access control;
- retention;

- consent;
- data lineage;
- privacy risk;
- third-party processing.

Development teams should include privacy requirements in user stories.

For example:

User Story: As a Data Principal, I want to understand how my personal data is used so that I can make an informed decision.

Acceptance criteria might include:

- clear notice displayed;
- purpose identified;
- relevant data categories listed;
- notice version recorded;
- consent status stored.

This converts legal principles into familiar software-engineering practices.

Implications for Lawyers

Legal professionals also need greater technological awareness.

A lawyer reviewing an AI application should be able to ask:

- Where is the data stored?
- What databases receive it?
- Is it copied into a training environment?
- Is an external AI API involved?
- Is the data pseudonymised?
- How is deletion implemented?
- What happens when consent is withdrawn?
- What happens to historical training data?
- Who has access to model outputs?

The future of technology law therefore requires lawyers to engage with architecture rather than only documentation.

Implications for AI Developers

AI developers should treat datasets as governed digital assets.

A dataset should have:

- provenance;
- owner;
- purpose;
- classification;
- legal status;
- retention period;
- permitted uses;
- security level;
- model associations.

This concept may be called a **Dataset Passport**.

Before model training, the dataset should pass through a compliance gate.

This would transform:

“Can we train this model with this dataset?”

into a formal technical and legal question.

Implications for Organisations

Organisations should move from annual or periodic privacy audits toward continuous compliance monitoring.

This resembles modern DevSecOps.

Traditional:

Develop → Deploy → Audit

Proposed:

Develop → Test Security → Test Privacy → Test Compliance → Deploy → Monitor → Update

The distinction becomes particularly important as AI systems continuously evolve.

Implications for India

The Indian context presents an important opportunity.

The DPDP Act provides the statutory foundation, while the 2025 Rules add operational detail and introduce provisions referring to techno-legal measures. ([MeitY](#))

India therefore has an opportunity to develop a distinctive privacy-engineering ecosystem.

Universities, technology companies, regulators and legal institutions could collaborate on:

- compliance software;
- privacy engineering education;
- AI governance standards;
- legal-tech tools;
- automated audit systems;
- privacy-preserving AI;
- regulatory sandboxes.

This would be especially valuable for India's rapidly expanding technology sector.

Implications for Education

The research also has implications for higher education.

Computer Applications programmes traditionally emphasise:

- programming;
- databases;
- networking;
- software engineering;
- web technologies;
- AI and machine learning.

Law programmes traditionally emphasise:

- legislation;
- constitutional law;
- jurisprudence;
- contracts;
- intellectual property;
- cyber law.

LCbD creates a natural interdisciplinary space between the two.

A university could establish a course titled:

Legal Technology, Privacy Engineering and AI Governance

Students could develop a prototype compliance platform as a practical project.

Such education would prepare graduates for emerging careers in:

- privacy engineering;
- legal technology;
- AI governance;
- data protection;
- cybersecurity compliance;
- technology law.

VI. Conclusion:

Conclusion

Artificial Intelligence has changed the scale and complexity of personal-data processing. The issue is no longer simply whether organisations collect information. Modern AI systems can infer, predict, classify, transform, reproduce, and share information across complex digital ecosystems.

This creates a fundamental challenge for law and technology.

Law establishes rights and obligations.

Software determines what happens in practice.

The central contribution of this paper is the proposition that these two dimensions should be connected through

Legal Compliance by Design.

The proposed LCbD framework extends Privacy by Design by establishing a chain between:

Legal Rule → Software Requirement → Technical Control → Verification → Audit Evidence → Continuous Monitoring.

The framework includes ten major components covering legal mapping, data inventory, consent, purpose limitation, retention, individual rights, AI dataset governance, security, auditability, and legal monitoring.

The analysis demonstrates that significant portions of privacy compliance can be supported through software. Data inventories can be automated. Retention can be monitored. Consent can be recorded. Data lineage can be tracked. Security controls can be tested. AI datasets can be governed. Audit evidence can be generated.

However, the study also establishes an equally important limitation: **law cannot be completely reduced to software rules.**

Legal interpretation often requires context, judgment, proportionality, institutional knowledge, and awareness of changing jurisprudence. Consequently, a responsible compliance architecture must maintain human oversight.

The most appropriate model is therefore not:

Human versus Machine

but:

Human + Machine.

Machines provide consistency, scale, speed and monitoring.

Humans provide interpretation, judgment, accountability and ethical reasoning.

The future of AI governance will require both.

Limitations

This study has several limitations.

First, the proposed framework is conceptual rather than a full-scale industrial implementation.

Second, the risk-scoring algorithms presented are illustrative. They should be calibrated against empirical organisational data before being used for regulatory or legal decision-making.

Third, the Indian data-protection framework remains relatively new, and organisations must monitor subsequent notifications, regulatory practice, judicial interpretation and implementation timelines.

Fourth, sector-specific laws may create additional obligations beyond the general data-protection framework.

Fifth, no software system can guarantee complete privacy.

Finally, legal compliance should not be confused with ethical responsibility. A system may technically satisfy minimum legal requirements while still creating social or moral concerns.

The principal conclusions can be summarised as follows:

1. **Privacy is increasingly a software-engineering issue.**
2. **AI creates privacy risks throughout the entire data and model lifecycle.**
3. **Legal requirements should be incorporated into software requirements engineering.**
4. **Data lineage is essential for AI compliance.**
5. **Privacy controls should be implemented as technical controls wherever possible.**
6. **Automated compliance can support, but cannot replace, human legal judgment.**
7. **Legal-technical traceability can strengthen accountability.**
8. **Cybersecurity and privacy should be designed together.**

9. **AI datasets should be governed as controlled digital assets.**
10. **India has an opportunity to develop an India-specific privacy-engineering and AI-governance ecosystem.**

The ultimate objective should not be to create software that merely says “**compliant.**”

The objective should be to create software that can explain:

What data are we using?

Why are we using it?

Who authorised the use?

Where is it stored?

Who can access it?

How long will we retain it?

What happens if the individual objects?

What happens if the law changes?

And where is the evidence that we did what the law requires?

That is the essence of Legal Compliance by Design.

References

1. Cavoukian, Ann. 2011. *Privacy by Design: The 7 Foundational Principles*. Information and Privacy Commissioner of Ontario.
2. European Union. 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016: General Data Protection Regulation*. Official Journal of the European Union.
3. European Union. 2024. *Regulation (EU) 2024/1689 of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*. Official Journal of the European Union.
4. Government of India. 2023. *Digital Personal Data Protection Act, 2023*. New Delhi: Ministry of Law and Justice.
5. Government of India. 2025. *Digital Personal Data Protection Rules, 2025*. New Delhi: Ministry of Electronics and Information Technology.
6. Kosenkov, O., E. Zabardast, D. Fucci, D. Mendez, and M. Unterkalmsteiner. 2026. “Privacy by Design: Aligning GDPR and Software Engineering Specifications with a Requirements Engineering Approach.” *Information and Software Technology* 190: 107946.
7. Nissenbaum, Helen. 2010. *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford University Press.
8. National Institute of Standards and Technology. 2020. *NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management*. Version 1.0. Gaithersburg, MD: U.S. Department of Commerce.
9. National Institute of Standards and Technology. 2023. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. Gaithersburg, MD: U.S. Department of Commerce.
10. National Institute of Standards and Technology. 2023. *NIST AI RMF Playbook*. Gaithersburg, MD: U.S. Department of Commerce.
11. National Institute of Standards and Technology. 2024. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. NIST AI 600-1. Gaithersburg, MD: U.S. Department of Commerce.
12. Nissenbaum, Helen. 2004. “Privacy as Contextual Integrity.” *Washington Law Review* 79 (1): 119–58.
13. Puttaswamy, K. S. (Retd.) v. Union of India. 2017. *Justice K. S. Puttaswamy (Retd.) and Another v. Union of India and Others*, (2017) 10 SCC 1.
14. Solove, Daniel J. 2006. “A Taxonomy of Privacy.” *University of Pennsylvania Law Review* 154 (3): 477–564.

15. Tabassi, Elham. 2023. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology.
16. Warren, Samuel D., and Louis D. Brandeis. 1890. "The Right to Privacy." *Harvard Law Review* 4 (5): 193–220.
17. Westin, Alan F. 1967. *Privacy and Freedom*. New York: Atheneum.

BIOGRAPHIES

1st Author 	A teaching career spanning over nine years has been shaped, where responsibilities such as Academic Co-ordinator, Nodal Teacher, Routine-in-Charge, and Governing Body Member were shouldered. Presently, working as Associate Professor along with duties as Exam Co-ordinator are being fulfilled at G H Raisoni University.	2nd Author 	An Assistant Professor whose research centers on artificial intelligence ethics, neural network applications, cloud cost optimization (FinOps), and IoT networking. He has authored peer-reviewed journal papers, contributed chapters to edited engineering volumes, and holds multiple registered design patents.
--	---	--	--