

Handbook on Data Protection and Privacy for Developers of Artificial Intelligence (AI) in India:

Practical Guidelines for Responsible Development of AI

July 2021



Published by:
Deutsche Gesellschaft für
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices
Bonn and Eschborn

FAIR Forward: Artificial Intelligence for All
A2/18, Safdarjung Enclave,
New Delhi - 110029, India
T: +91 11 4949 5353
F: + 91 11 4949 5391

E: info@giz.de
I: www.giz.de

Responsible:
Mr. Gaurav Sharma
Advisor – Artificial Intelligence
FAIR Forward: Artificial Intelligence for All
GIZ India
Gaurav.sharma1@giz.de

Authors:
KOAN Advisory (Varun Ramdas, Priyesh Mishra, Aditi Chaturvedi)
Digital India Foundation (Nipun Jain)

Content Review:
Gaurav Sharma, GIZ India

Editor:
Divya Joy

Design and Layout:
Caps & Shells Creatives Pvt Ltd.

On behalf of
German Federal Ministry for Economic Cooperation and Development (BMZ)

GIZ India is responsible for the content of this publication.

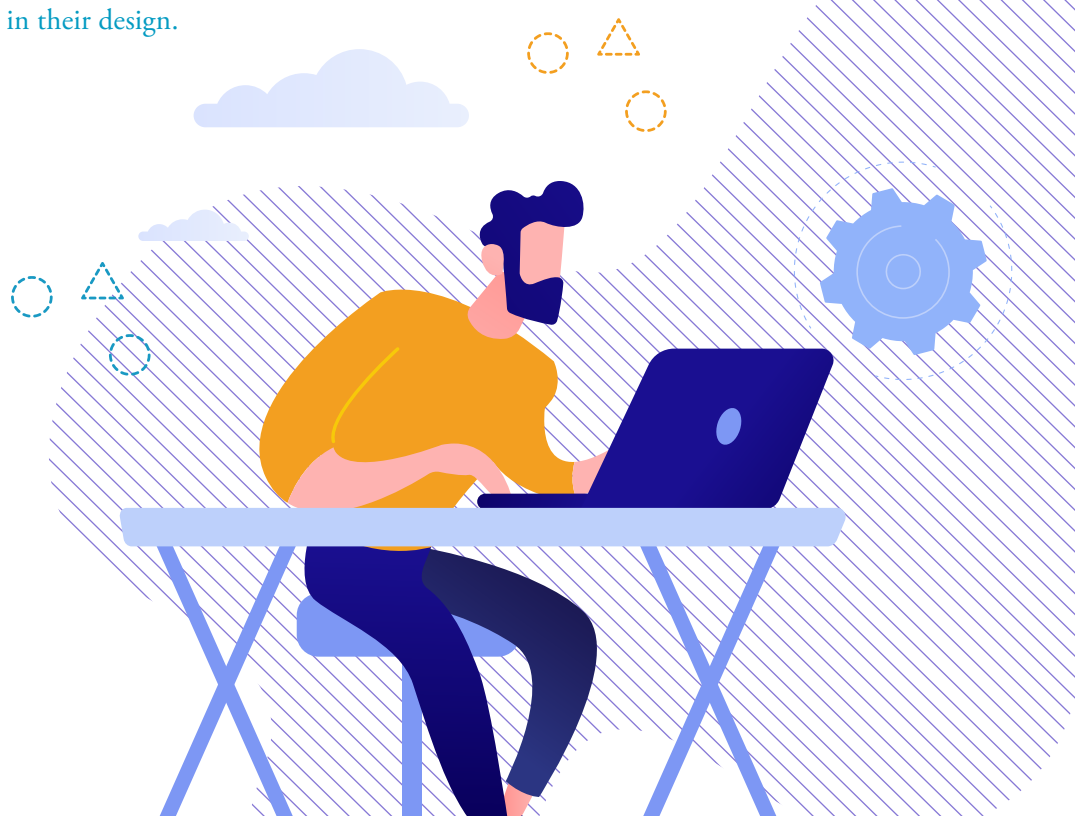
Disclaimer: The data in the publication has been collected, analysed and compiled with due care; and has been prepared in good faith based on information available at the date of publication without any independent verification. However, Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH does not guarantee the accuracy, reliability, completeness or currency of the information in this publication. GIZ shall not be held liable for any loss, damage, cost or expense incurred or arising by reason of any person using or relying on information in this publication.

New Delhi, India
July 2021

What to expect from the handbook and prerequisites?

Trust is the key to widespread adoption and success of Artificial Intelligence (AI). It is for this reason that AI products must adhere to the highest standards of ethics and prevailing regulations. This data protection handbook shall act as a guiding document and help in the development of artificial intelligence technologies that are legally and socially acceptable.

This handbook is for developers of AI who are already familiar and know Machine Learning (ML) processes such as early-stage start-ups. The purpose of the handbook is to give developers a clear understanding about the essential ethical frameworks in AI and important issues related to protection of personal data and privacy. With the help of this handbook, developers will be able to follow the guidelines on ethics and data protection laws from the very inception of product/application design to its development. Keeping these parameters in mind from the beginning can save developers and start-ups from issues that may arise later on and can create complications that may even force abandoning and redesigning product/application. The handbook encourages developers to take an interdisciplinary view and think beyond the confines of algorithmic accuracy and focus on the social impact of the product. The handbook does not provide technical solutions but gives developers ethical and legal objectives to pursue. Developers are free to think of innovative ways in which they can include these guidelines in their design.



Organisations involved



Koan Advisory is a New Delhi-based public policy consultancy firm. It has a multi-disciplinary team of lawyers, economists, social scientists and communications professionals who work with clients in the public and private sector to help shape policy discourse in India. Koan Advisory's subject area expertise includes intellectual property and innovation, competition and market structures, trade and commerce, and governance of new and emerging technology.



Digital India Foundation is a policy think-tank promoting Digital Inclusion, Cyber Security, Mobile Manufacturing, Domestic Consumption, Software Products and Smart Cities. Since its inception in 2015, it has worked with pioneers in industry, academia and government organizations on issues ranging from promoting a safer and inclusive product and policy environment for Indian ecosystem.



Data Security Council of India (DSCI) is a not-for-profit, industry body on data protection in India, setup by NASSCOM®, committed towards making cyberspace safe, secure and trusted by establishing best practices, standards and initiatives in cyber security and privacy. DSCI works together with the Government and their agencies, law enforcement agencies, industry sectors including IT-BPM, BFSI, CII, Telecom, industry associations, data protection authorities and think tanks for public advocacy, thought leadership, capacity building and outreach initiatives. DSCI through its Centres of Excellence at National and State levels, works towards developing an ecosystem of Cyber Security technology development, product entrepreneurship, market adoption and promoting India as a hub for Cyber Security.



Artificial Intelligence for all.



The German Development Cooperation initiative "FAIR Forward – Artificial Intelligence for All" strives for a more open, inclusive and sustainable approach to AI on an international level. To achieve this, we are working together with five partner countries: Ghana, Rwanda, South Africa, Uganda and India. Together, we pursue three main goals: a) Strengthen Local Technical Know-How on AI, b) Remove Entry Barriers to AI, c) Develop Policy Frameworks Ready for AI.

Foreword



Artificial Intelligence (AI) has emerged as a foundational technology, powering NextGen solutions across a wide cross-section of verticals and use cases. The innovation ecosystem building AI based products and solutions for BFSI, healthcare, retail, cyber security to name a few, spans the Big Tech, start-ups, and services firms. With the global AI market size estimated to grow from USD 37.5 bn in 2019 to USD 97.9 bn by 2023, the growing demand for talent pool and their skills upgrade is a priority for India's technology industry.

Both governments and businesses are favoring AI not only for its innovation potential but also for revenue impact and cost savings. As a result, it becomes increasingly relevant to mitigate the risks of AI solutions on users and society at large. In the context of AI, cases of algorithmic bias or compromised safety come to the forefront with the risk of undoing the potential benefits of leveraging this technology for good and creating wider socio-economic impact. Awareness and capacity building of the developer community is essential for creating a strong foundation for responsible data-centric innovation. This is an important cog in the overall schema for the proliferation of AI technology in India and balancing its innovation potential with ethics and privacy.

It is imperative that risks of abuse of AI technology are not only addressed through inclusive policymaking but also practitioner-centric initiatives to integrate best practices of ethics and privacy into the AI solutions proactively. DSCI has always championed the practitioner-led approach as a means of making a tangible impact in the policy discourse. Educating the developers and solution architects innovating on AI on all the right practices to meet both user and regulator expectations will create a longer-term positive impact.

To facilitate this, the Data Security Council of India (DSCI) is delighted to collaborate with the German Development Cooperation (GIZ), Digital India Foundation, and Koan Advisory Group to come out with this handbook, to assist developers in building AI solutions that take into account ethical and privacy considerations. The handbook draws from globally recognized ethical principles and the current and evolving regulatory landscape for privacy in India. We are confident the handbook will help developers follow the guidelines on ethics and data protection laws from the very inception, of product/application design to development. The handbook encourages developers to take an interdisciplinary view, think beyond the confines of algorithmic accuracy and give a thought to the social impact of the product and adequate attention to accountability and transparency, right from the design stage. Several experts from industry and academia contributed with their rich experiences and we thank them for making this handbook of immense value add to the AI ecosystem and motivate the developer talent pool on a journey of discovering responsible AI innovation.



Rama Vedashree
CEO
Data Security Council of India

Foreword

The world is advancing towards an Artificial Intelligence (AI)-powered industrial revolution. However, the rapid development and deployment of this emerging technology is accompanied by significant uncertainties that may erode public trust in AI. For instance, tough questions about the liabilities of AI systems and the risks posed by algorithmic biases, need answers.

Though such complex challenges are hard to solve from a rulemaking perspective alone, they can be overcome if appropriate ethical considerations shape two key aspects of AI development: the use of data – the building block of AI – and the deployment of algorithms that are trained on such data.

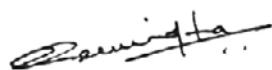
Software developers are the architects of AI systems and they must be the first to understand the ethical considerations and emerging legal frameworks that govern them. An awareness among the developer community is arguably more important than any other facet of AI adoption at scale. For example, greater sensitivity to AI ethics and laws can help developers identify sensitive data and build safeguards to manage risks linked to it.

This handbook, co-developed by the Digital India Foundation and the Koan Advisory Group, with technical and financial support from the German Development Cooperation (GIZ) and the Data Security Council of India (DSCI), is meant to be an easy reference guide that helps developers put into practice ethical and legal requirements of AI systems.

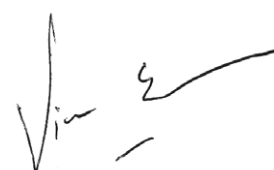
The handbook discusses key concepts like bias, privacy, data security, transparency and accountability. The best practices, checklists and citations mentioned here are meant to facilitate a rigorous yet, pragmatic and actionable understanding of the subject at hand.

The ethical principles mentioned in the handbook are in line with the principles proposed by the Organisation for Economic Co-operation and Development and India's NITI Aayog. The legal-regulatory requirements correspond to the provisions under India's Draft Data Protection Bill, 2019 and Information Technology Act, 2000.

A fearless embrace of AI requires supportive baseline conditions. This means that society at large must be convinced of the symbiotic nature of technology, through participative and robust design. We have kept these inclusive first-principles in mind to create this instructional guide. We thank all the developers and domain experts who helped us shape this handbook and hope that its contents are widely discussed among relevant communities.



Arvind Gupta
Head & Co-Founder
Digital India Foundation



Vivan Sharan
Partner
Koan Advisory

Acknowledgement



The Handbook on Data Protection and Privacy for Developers of Artificial Intelligence in India: Practical Guidelines for the Responsible Development of AI is a comprehensive document which details the rules and best practices on the ethical use of AI. Multiple organizations collaborated with each other to develop a handbook that will work as a guide for developers to develop such software tools that adheres to the ethical and moral principles of privacy, dignity, fairness and equity in the use of AI by different entities.

It gives me immense pleasure to express my heartfelt gratitude to all the organisations and individuals who made the writing of this handbook a learning experience. This document is the result of an intense deliberation of a small but diverse group of stakeholders. These include (in alphabetic order); Amith Parameshwara, Leader - Artificial Intelligence at Kimberly-Clark; Amit Shukla, Founder, EasyGov; Anand Naik, CEO & Co-Founder at Sequaretek; Atul Khatavkar, CEO HB Infotech; Bhavesh Manglani, Co-Founder at Delhivery; Professor Biplav Srivastava, Computer Science and Engineering, AI Institute, University of South Carolina (USC); Dr Kalika Bali, Principal Researcher at Microsoft Research India (MSR); Girish Nair, Founder, Curiosity Gym; Dr Neeta Verma, Director General of National Informatics Centre; Saurabh Chandra, Co-Founder, Ati Motors; Prof Siddharth Shekhar Singh, Associate Dean - Digital Transformation, e-Learning, and Marketing at Indian School of Business; Shuchita Gupta, Co-Founder, Care4Parents; UmakantSoni, Co-founder & CEO AI and Robotics Technology Park (ARTPARK); Professor Venkatesh Balasubramaniam, Department of Engineering Design, IIT Madras; Vibhore Kumar, CEO & Co-founder at Unscrambl Inc.

We also extend our thanks to Philipp Olbrich, Gaurav Sharma & Siddharth Nandan from FAIR Forward: AI for All project, GIZ; Anand Krishnan from Data Security Council of India; Aditi Chaturvedi, Priyesh Mishra, Varun Ramdas and Vivan Sharan from Koan Advisory as well as Arvind Gupta and Nipun Jain from Digital India Foundation. Without the guidance of these individuals and constant mentorship from Professor Biplav Srivastava, the handbook wouldn't have been in the form it is today.

Thank you!

Contents

10	How to read this Handbook?	36	V. Security
10	Meaning – Brief explanation of principle and why it is important	36	Meaning
11	Section I : Ethics in AI	37	Checklist for Developers at Different Points of Intervention
12	Introduction	38	Good Practices
13	Stages at which developers can intervene	39	At a Glance
15	Ethical Principles	39	Instances of challenges developers may face
15	I. Transparency	40	Further Questions
15	Meaning	41	VI. Privacy
16	Checklist for Developers at Different Points of Intervention	41	Meaning
17	Good Practices	42	Checklist for Developers at Different Points of Intervention
18	At a Glance	43	Good Practices
18	Instances of challenges developers may face	44	At a Glance
19	Further Questions	44	Instances of challenges developers may face
20	II. Accountability	45	Further Questions
20	Meaning	45	Ethics in AI literature and policy documents
21	Checklist for Developers at Different Points of Intervention	46	Section II: Data Protection
22	Good Practices	47	Introduction
23	At a Glance	48	Data Protection under Information Technology Act, 2000
23	Instances of challenges developers may face	49	Sections under IT Act
24	Further Questions	49	I. Liability of corporate entities
25	III. Mitigating Bias	49	II. Reasonable security practices and procedures and sensitive personal data or information Rules, 2011
25	Meaning	49	i. Sensitive personal information
26	Some Reasons for Bias in AI	49	ii. Privacy policy
27	Checklist for Developers at Different Points of Intervention	50	iii. Collection of Information
28	Good Practices	50	iv. Disclosure of Information
29	At a Glance	50	v. Reasonable Security Practices and Procedures
29	Instances of challenges developers may face	51	Data Protection under Personal Data Protection Bill, 2019
30	Further Questions	51	Glossary
31	IV. Fairness	51	Personal Data
31	Meaning	51	Data Principal
32	Checklist for Developers at Different Points of Intervention	51	Data Fiduciary
33	Good Practices	51	Data Processor
34	At a Glance	51	Processing
35	Instances of challenges developers may face	51	Significant data fiduciary
35	Further Questions	51	Sensitive Personal Data

52	Principles of Personal Data Protection	73	Storage Limitation
54	Personal Data	73	At a glance
54	At a glance	73	What is the storage limitation principle?
54	Introduction	74	For how long can personal data be stored?
55	What is personal data?	74	When should you review the necessity of retention?
56	What is meant by direct or indirect identification?	75	When can personal data be stored for a longer period of time?
57	What should you do if you are unable to determine whether data is personal or not?		
58	Are there other categories of personal data?		
59	Data Fiduciaries and Data Processors	76	Data Quality
59	At a glance	76	At a glance
59	Introduction	76	What is the principle of data quality?
60	Who is a Data Fiduciary?	77	When is data accurate/inaccurate?
60	Who is a Data Processor?	77	Whose responsibility is it to maintain data quality?
61	How to determine if you are a data fiduciary or a data processor?		
62	Notice and Consent	78	Accountability and Transparency
62	At a glance	78	At a glance
62	What is the principle of notice and consent?	78	Introduction
63	How do you implement the principle?	79	Privacy by Design
63	What is the standard of valid consent?	80	How would Privacy by Design be implemented by law?
64	Are there exceptions to the principle of notice and consent?	81	What are accountability measures prescribed by law?
		82	What are the transparency requirements prescribed by law?
65	Purpose Limitation	84	Rights of Data Principals
65	At a glance	85	Right to confirmation and access
65	What is the meaning of purpose limitation?	86	Right to Correction and Erasure
66	How do you specify your purpose?	87	Right to Portability
66	Can the data be used for purposes other than the ones specified?	88	Right to be Forgotten
68	Data Minimisation	90	Compliance Map
68	At a glance	91	Annexure A
68	What is the principle of data minimisation?		Technical Measures for Data Security
69	How do you determine what is relevant and necessary?	93	Annexure B
70	Is data minimisation principle antithetical to AI?		List of Abbreviations
71	Are there legal provisions protecting big-data applications?	94	Annexure C
71	What data minimisation techniques are available for AI systems?		Master Checklists
		100	References

How To Read This Handbook?

The Handbook is divided into two sections - Ethics (Section I) and the Law (Section II). Section I covers Ethics in AI and Section II investigates the legal requirements for data protection.

Section-I

Based on existing AI literature and policy documents [Page 45], we discuss six ethical principles in Section I of this document.

1. **Transparency**
2. **Accountability**
3. **Bias**
4. **Fairness**
5. **Security**
6. **Privacy**

Under each section, the principles are explained in the following manner:

- Meaning
 - Checklist
 - Good Practices
 - At a Glance
 - Challenges
-
- **Meaning:** Brief explanation of the principle and why it is important.
 - **Checklist:** A quick reference checklist for developers at each stage of intervention: Pre-processing; Processing and Post-processing [explained on Page 13]
 - **Good Practices:** A list of good practices and frameworks sourced from existing literature
 - **At a glance:** Summary of the principle and quick tips
 - **Challenges:** Challenges that developers face. This is based on discussions with developers and examples borrowed from existing literature.

Section-II

Section II explains the data protection framework in India. The section is divided into two broad chapters:

1. **Data Protection under Information Technology Act, 2000**
2. **Data Protection under Personal Data Protection Bill, 2019**

The second chapter i.e., Data Protection under Personal Data Protection Bill, 2019 is divided into smaller chapters based on the principle of data protection it refers to. These are:

- Personal Data
- Data Fiduciary and Processor
- Notice and Consent
- Purpose Limitation
- Data Minimisation
- Storage Limitation
- Data Quality
- Accountability and Transparency
- Rights of Data Principal

Each chapter is structured in an FAQ format, with an 'At a glance' section summarising key points. Actionable items under each chapter are captured under 'Call to Action' boxes.

Each section also has a quick reference 'Checklist for Developers' that briefly summarises compliance requirements in a tickbox format".

A Compliance Map at the end of Section II captures the workflow and summarises actionable compliance requirements in form of a Master checklist.

Section I

Ethics in A I

“

In our way of working, we attach a great deal of importance to humility and honesty; With respect for human values, we promise to serve with integrity.

Azim Premji

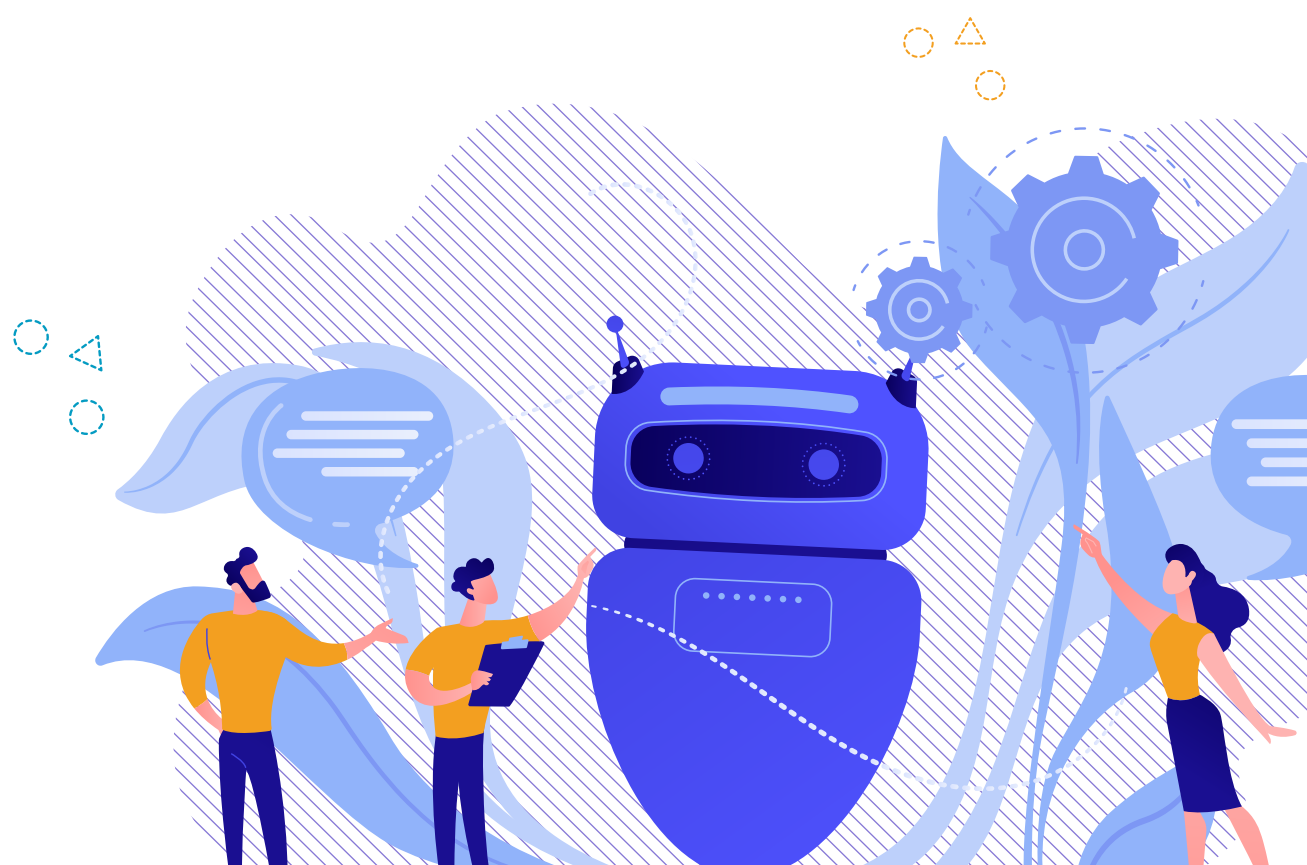
”

“

Technology alone is not enough. It's technology married with liberal arts, married with the humanities, that yields the results that make our hearts sing.

Steve Jobs

”



Introduction

Development in AI raises significant questions on law and society. The law answers some of these questions but cannot anticipate every technological breakthrough, given the speed of technological progress. Consequently, compliance with the law is mandatory but inadequate to meet all the concerns of technology and society. National governments, industry leaders and academia seek to fill this gap by promoting principles or frameworks for ethical AI.

The literal meaning of ethics is self-evident but there is want of a common understanding of 'ethics'. According to a report by the Alan Turing Institute on Understanding Artificial Ethics and Safety, "AI ethics is a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies." Establishing an ethical framework for AI starts with explaining the risks and opportunities regarding the design and use of technology. The objective of ethics in AI is

not to limit its scope and underuse technology, but to create an ecosystem where opportunities are maximised, and risk minimised. For developers and start-ups that deploy AI, an ethical approach provides a dual advantage. It enables them to harness social value by using new opportunities that society prefers and appreciates and helps them navigate potential risks.

An ethical framework for AI development is a governance structure that we can break down into design goals at each stage of development. In a nutshell, an ethical framework should consider:

- The potential impact of the project on stakeholders and communities.
- Anticipating discriminatory effects of AI project on individuals and social groups to assess if the project is fair and non-discriminatory
- Identifying and minimising biases in the dataset that may influence the model's outputs. Developers should

try to perform this check at every stage of design.

- Enhancing the level of public trust in the project through guarantees for transparency, accuracy, explainability, reliability, safety, and security on a best-efforts basis.
- Methods to explain the AI project to stakeholders and communities including the decision-making process and redressal mechanisms in place, to the extent possible.

The objective of ethics in AI is not to limit its scope and underuse technology, but to create an ecosystem where opportunities are maximised, and risk minimised.

Stages at which Developers can Intervene

Artificial Intelligence enabled systems contain different points of intervention to mitigate ethical concerns such as bias, incorrect/excessive training data, incorrect training algorithm, inadequate redressal. Based on our discussions with established start-ups, practitioners and developers, we have classified AI development into three stages for intervening in the decision-making process, from an ethical point of view. These are pre-processing, in-processing, and post-processing.

Pre-processing



“Pre-processing techniques seek to approach the problem by removing underlying discrimination from data prior to any modelling.” At the pre-processing stage, developers may take several steps to mitigate ethical concerns. For example, they may balance the datasets and enable data traceability. It is important to trace the lineage of data collected for training models and ensure that it is collected and stored following data protection regulations. At this stage, developers can also anticipate security and privacy risks with the model and prepare methods to mitigate these harms.

In-processing

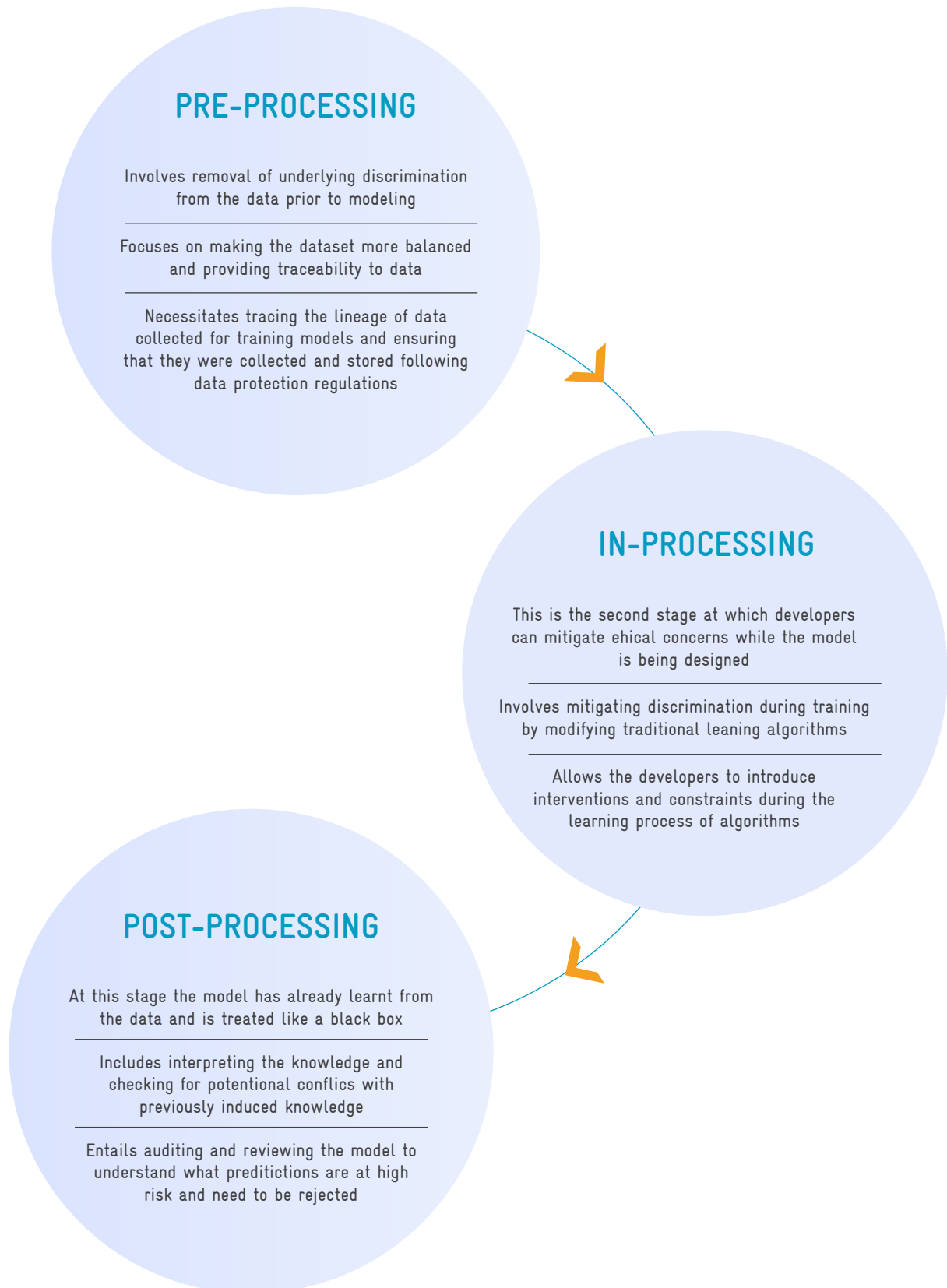


This is the second stage of the development process. At this point, developers can mitigate ethical concerns while the model is being designed. “In-processing techniques mitigate discrimination during training through modifications to traditional learning algorithms. At this stage, developers can intervene to encourage or restrict reliance on certain attributes during training, trace and identify entry points for intervention, and secure the model from security risks. It is also important to embed measures to allow diagnostics and visibility of the training process.

Post-processing

This refers to a stage where the AI model has already learnt from the data. Complex AI models are constructed as a black box. Lack of visibility into AI functioning is detrimental to transparency and accountability and prevents any course correction. For post processing assessment, it is important to build in checks and balances. Post-processing techniques include interpreting the knowledge and checking for potential conflicts with previously induced knowledge. Post-processing audit and review mechanisms are critical to inculcate explainability and hence public trust in AI models.

Figure 1: Stages of Intervention



Ethical Principles

I. Transparency

Meaning

Complex algorithmic systems are commonly called ‘black boxes’ because you can see the inputs and outputs, but not the process behind the results. Transparency is a design choice that can help see into this black box, increase accountability, and enhance public trust in AI decision-making.

Transparency provides insight into a system, and is the first step towards making AI explainable, redressable, appealable and accountable.

There are multiple levels of transparency. For a user/beneficiary, transparency means an easily understandable explanation of the decision-making process followed by the AI. For developers, it means source code visibility. Transparent design informs users that AI systems are making automated decisions and helps users understand the basis for these decisions. This empowers users with the capacity to dispute or appeal it.

For a developer, transparency refers to the ability to read, interpret and understand underlying design supporting an AI-enabled system. From a practical perspective, transparency means a developer should be able to diagnose errors in decision making, find intervention points and correct automated decisions.

Scenario:

A loan applicant comes to know that her bank has denied her loan application to start a new business. The bank tells her that their AI enabled system considered her application and rejected her loan. She has no knowledge of AI and responds that she has never defaulted on bills and this is the first time she is applying for a loan. The bank employee tells her that he does not understand the decision-making either.

How will transparency help?

Transparent design allows loan applicants to contest automated decisions. The AI enabled system may have based the decision on incorrect training data or by considering wrong factors. For example, the AI could have considered attributes like age, gender, or marital status and made the decision in the above case. Transparency should allow the bank and the applicant to revisit the attributes considered by the AI system. From the developer’s perspective, transparency is important to review the weightage given by the system to each attribute and course correct, as necessary.



Checklist for Developers at Different Points of Intervention

Pre-Processing

Are you aware of the source of data used for training?

Have you recorded the attributes to be used for training and the weightage for each?

Is there scope for diagnosing errors at a later stage of processing?

In-processing

Have you checked for any blind spots in the AI-enabled system and decision-making process?

Can the AI decision-making process be explained in simple terms that a layperson can understand?

Post-processing

Is it possible to obtain an explanation/reasoning of AI decisions? For instance, can a banker obtain a record for acceptance/denial of a loan application.

Is there a mechanism for users and beneficiaries to raise a ticket for AI decisions?

Is there scope for oversight and human review of AI decisions?



Good Practices

- Transparency can be achieved through hygiene governance. Developers can build (i) operational and development protocols to ensure design review and explainability; (ii) clear and meaningful articulation of terms of use; (iii) transparency impact assessments and periodic review.
- The Defense Advanced Research Projects Agency created a model for [Explainable AI \(XAI\)](#) that makes automated decisions explainable. It seeks to provide an explanation for AI outcomes. Local Interpretable Model Agnostic Explanations ([LIME](#)) is one such approach.
- LIME is a model-agnostic solution that provides explanations anyone can understand. LIME is an algorithm, which can explain the predictions of a classifier. It can help determine if one can trust (a) a predication sufficiently to take action based on it and (b) trust the model enough for it to behave in reasonable ways. Some desirable characteristics of explanation methods are – interpretability i.e. provide qualitative understanding between input variable and response.
- Transparency by Design is a practical guide to help promote the benefits of transparency while mitigating challenges posed by AI enabled systems. [Researchers](#) have developed a model that could aid start-ups design transparent systems, based on a set of nine principles that incorporate technical, contextual, informational, and stakeholder-sensitive considerations. It also supplies a step-by-step roadmap on integrating these principles in the pre-processing stages,
- Another solution to improve transparency is a [decision tree](#). All data is displayed in the form of a tree. A developer can use this as a diagnostic tool by starting at the top of each level and selecting a branch based on the value of a particular feature. The outcome at the last leg or the foundation is the outcome. This simple model can provide a high degree of transparency for the developer, in instances where the model is based on a manageable amount of data.
- Transparent models such as [Rule Based Learners](#), [General Additive Models](#), [Bayesian Models](#), [K-Nearest Neighbours](#) can supplement the decision tree model based on the complexity of variables, levels involved and how the AI interacts with the variables.
- Post-hoc explainability is another model. In this model, the nature and properties of outputs yielded by the black box can be reverse engineered to explain outcomes. However, this need not be accurate all the time.



At a Glance

To ensure transparency, developers may consider developing the following protocols.

An easy-to-explain model of the decision-making process (like a decision tree).

List of attributes considered and weightage for each.

Map of intervention points.

Redressal mechanism for users and beneficiaries.

Instances of challenges developers may face

Devices with non-visual interfaces¹: Developers may need to think of ways in which transparency requirements can be presented to users when they develop products such as smart speakers (as opposed to services like dealing with loan applications), where it may not be feasible to explain these issues through visual means.

¹ For example Alexa

Further Questions

Is transparency antithetical to Intellectual Property Rights (IPR)?

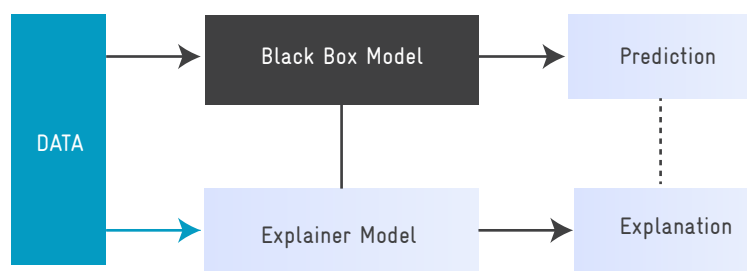
Transparency has to be balanced against intellectual property protection. Algorithmic systems involve significant investment by private firms and often they are protected as trade secrets or copyright. Hence, transparency does not refer to full disclosure of all aspects of the algorithm or the source code supporting it.

Why is transparency important?

Some algorithmic systems are so complex that they are often called 'black boxes' because you can see the inputs and outputs, but not their exact functioning/processing. Although it is challenging to understand how automated systems work, the proposed checklist for transparency can enhance trust in systems and improve overall efficiency of AI. Transparency can ensure that AI systems do not use data in a way that leads to biased outcomes. It also ensures better compliance with the laws and enforces legal rights. The ability to explain the processes involved can aid developers identify the vulnerabilities involved.

The US Center for Tropical Agriculture data-driven agronomy workintentionally designs models for ease of interpretation. This helps them gain insights on the relationship between crop management practices and crop yield. If they had modelled their system using "black box" algorithms, they would be able to predict the yield but would have no visibility on the specific inputs that most strongly determined the yield.

Figure 2: Explainer Model



Flowchart showing that the Explainer model takes the data and the prediction values as the input and creates a explanation based on shapely values. Source - PyCon Sweden : Interpreting ML Models Using SHAP by Ravi Singh



II. Accountability

Meaning

AI is non-human, and it is impossible to hold AI accountable for its decisions within the current legal framework. However, developers can identify intervention points at the stage of algorithm design, AI system training and data collection. Accountability in AI decision-making means it is possible to trace AI outcomes/decisions to an individual or an organization or a step in the AI design process.

Intuitively, the designer or the

start-up could be held responsible, but the challenge is that decision-making by AI systems undergo significant changes as they scale-up. Another reason why it is difficult to attribute responsibility to AI actions is that there may be multiple actors undertaking the various activities. For example, while one actor may design the system, other may train the system using data collected by someone else, and another would make it available for users/beneficiaries. This is

known as the “many hands problem” which is associated with layered handling and decision making at different stages of complex systems.

The purpose of accountable design is to both provide users/beneficiaries with a redressal mechanism and to solve problems internally as they come up. As there is no legal clarity on holding AI accountable, accountability as an ethical principle is the obligation to enable redressal and address design flaws.

Scenario:

A loan applicant comes to know that her bank has denied her loan application to start a new business. The bank tells her that their AI enabled system considered her application and rejected her loan. She wants to appeal the decision, but the banker is unable to guide her to the appropriate authority.

How will accountability help?

Accountable design and process would help the user seek an explanation on why she was denied a loan and seek a review of the decision. Accountability would also help the banker maintain its customer relations by guiding customers to the appropriate redressal point. In an ideal scenario, the bank files for a decision review and the AI provides the reasoning or factors considered for an outcome. The bank may then analyse this and reverse or maintain the decision to deny a loan.



Checklist for Developers at Different Points of Intervention

Pre-Processing

Does the start-up have policies or protocols or contracts on liability limitations and indemnity? Are these accessible and clear to you?

Does your organisation have a data protection policy? Are data protection guidelines being followed in the collection and storage of data, if the data is procured from a third-party?

Are there adequate mechanisms in place to flag concerns with data protection practices?

In-processing

Can the AI-enabled system be compartmentalised based on who develops the specific field?

Is it possible to apportion responsibility within your set-up if multiple developers have worked on a project?

Is it possible to maintain records on design processes and decision-making points?

Post-processing

Can decisions be scrutinised and traced back to attributes used and developers who worked on the project?

Have you identified tasks that are too sensitive to be delegated to AI systems?

Are there protocols/procedures in place for monitoring, maintenance and review of AI enabled systems?

Good Practices

- Establish an auditing mechanism to identify unwanted consequences, such as unfair bias, a solidarity mechanism to deal with severe risks in AI intensive sectors.
- A redressal mechanism to remedy or compensate for wrongs and grievances caused by AI. A widely accessible and reliable redressal mechanism is necessary to build public trust in AI. A clear and comprehensive allocation of accountability to humans and/or organisations is paramount in such a scenario.
- A good example is the ‘[Exception Handling Measures](#)’ provided by the Unique Identification Authority of India. It lists Application Programming Interface (API) errors and guidelines for handling errors within the application at different stages of authentication and service delivery.
- Accountability is the idea of tracing responsibility to human beings instead of saying that technology is responsible. For this, it is important to identify decision making processes that are too sensitive to be delegated to AI systems. Some accountability concepts are:
 1. “human in the loop” - ensuring that there is human intervention in cases where fully automated decisions are made
 2. “human over the loop” - human involvement in adjusting parameters.

[NITI Aayog Accountability Framework](#)

The NITI Aayog suggests that a framework for identifying failure points in AI could include the following. A legal framework for accountable AI could borrow from this framework. This framework indicates the extent of liability that could be built into such systems in future.

- A negligence test for damages caused by AI enabled systems, instead of strict liability. This can be achieved through self-regulation by conducting damage impact assessment at each stage of development.
- Safe harbours may be formulated to insulate or limit liability if proper steps are taken by the individual to design, test, monitor, and improve the system.
- A policy on accountability may include actual harm requirements so that lawsuits cannot be filed based on speculative damages.

[Algorithmic Impact Assessment by AI Now Institute](#)

The AI Now Institute suggests a practical accountability framework that involves public input and participation in assessing the impact of an AI enabled system on people. The Institute developed this framework based on existing impact assessment frameworks for human rights, environmental protection, data protection, and privacy. Although the mechanism is proposed for public agencies mainly, developers could also learn from this framework. Key elements of the assessment include:

- An internal assessment of existing systems and systems under development. An evaluation of potential impact on fairness, justice, bias, or other concerns of the user/beneficiary community.
- Meaningful review processes by an external/third-party researcher to track the progress of the system and its impact.
- A public notice with an easy-to-understand explanation of the decision-making process followed by the AI, and information on the internal assessment and third-party review process mentioned above.
- Public consultation with affected users/beneficiaries/stakeholders to address concerns and clear doubts.
- A mechanism for the public to raise concerns on the above process or any other substantive concerns.



At a Glance

Enhancing accountability requires a strategy that draws a clear chain of responsibility depending on the harm or failure caused by the AI system. This provides a path to identify where or who in the process is responsible for specific actions by AI systems and subsequently provide a redressal route. To ensure accountability, developers should consider developing the following protocols.

A diagnostic tool to verify that data was collected legally.

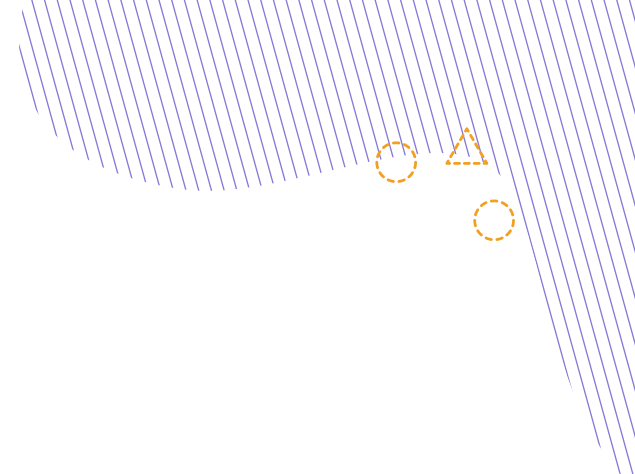
A map of the system and individual responsible for each function.

Redressal mechanism for beneficiaries of the system.

Instances of challenges developers may face

There are certain AI technologies where humans may not have adequate control or entry points for intervention. Autonomous vehicles are one such example. Often the technology may be designed to respond much quicker than humans. In such a scenario, it is difficult to intervene before an accident. One can only diagnose an error after.

Another concern is the problem of many hands or distributed responsibility. When multiple developers are involved in the development of a technology and in cases where the AI learns from external output, it is difficult to pinpoint one problem area. In 2016, Microsoft launched a chatterbox, Tay that produced misogynistic sentences and racist slurs after a point. Designers of the chat box were held responsible, but there was no accountability on the part of the users who may have introduced the bot to racist and misogynistic words.



Further Questions

What is the difference between transparency and accountability?

Transparency and accountability are two closely related concepts. The difference is that transparency is simply visibility on the functioning of an AI enabled system and accountability is the ability to trace each function to an individual.

Transparency is important to review systems internally while responsibility is assigned mostly for external action.

Why is accountability important?

If AI systems function without any accountability, there is no obligation to ensure that decision-making adheres to legal and/or ethical principles. This could lead to adverse outcomes such as service denial. In projects such as Aadhar², a biometric failure could lead to denial of access to government ration, cooking gas, bank accounts or other government support. The problem is further exacerbated if these people do not have a point of contact to address concerns. The need here is to compartmentalise AI systems so that errors can be traced back to design and effectively redressed.

² Aadhar is a 12-digit unique identity number issued by the Unique Identification Authority of India (UIDAI) for identity verification and transfer of government benefits.



III. Mitigating Bias

Meaning

In the context of automated decision-making by AI-enabled systems, we can understand bias to mean outcomes that are structurally less favourable to individuals from a certain group without a rationale justifying this distinction. Without checks and balances, bias amplifies within an AI-enabled system and leads to discriminatory outcomes on a group, even though a developer or designer may not have intended this. There is no legal framework to mitigate these harms. So, it is important for developers to voluntarily trace,

treat and mitigate these harms ethically to prevent unintended outcomes.

AI systems reproduce socially embedded or historical discrimination. This could be a design issue or an issue with the training data used or even a subsequent learning of the AI-enabled system once it starts learning from the public. As AI systems scale up, the social impact of these biases also grows.

Bias in AI may show up due to several reasons such as unrepresentative training data,

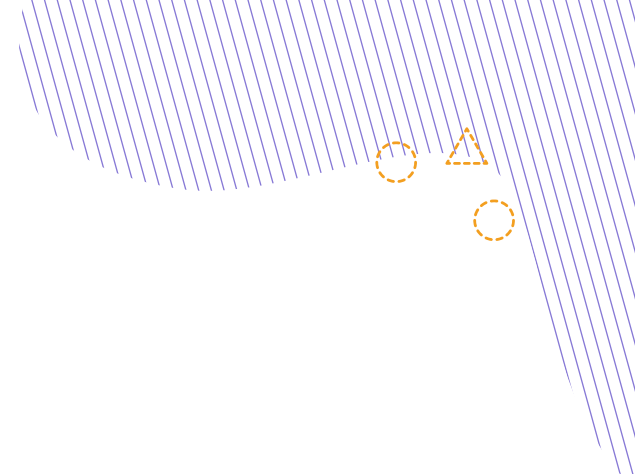
unconscious assumptions of programmers or biases in the community norms. It could also be a result of the technical architecture, data on which the system is built, or the context in which the system is deployed. Bias can stem from using attributes such as gender, race, or caste as training data without checks, and in some cases, without tracking how the AI system is processing data. Bias significantly diminishes the accuracy of an AI system.

Scenario:

A loan applicant comes to know that her bank has denied a loan application to start a new business. A male friend of hers with the same credit rating and similar economic background is granted the loan. The bank tells her that their AI enabled system considered her application and rejected her loan. On further investigation, it is revealed that in its 100 years of existence, the percentage of loans granted to unmarried women was less than 1 per cent. This training data was fed into the AI enabled system.

How do you mitigate bias in this scenario?

The first step here is to trace the source of bias. In most instances, bias in AI is an amplification of biased training data or biased design. Once the source has been identified, multiple practices exist to mitigate this bias. It is important to be cognizant of structural biases that exist in society to be able to easily identify the reason behind biased outcomes. At an organisational level, the start-up could adopt equal opportunity policies, recruitment to ensure diversity and representation, and awareness initiatives to educate developers/designers on social realities.



Some Reasons for Bias in AI

Reasons for bias	Explanation
Insufficient data collection	Data collected may be insufficient to represent the social realities of the space that the AI targets. Due to this, AI may not be able to attain its desired output.
Insufficient diversity in data	<p>Data may not be sufficiently diverse to capture all facets of the group an AI-enabled system seeks to work for. In such cases, the data might end up training the AI to discriminate against under-represented groups.</p> <p>For instance, an AI to detect cancer and trained on data available in North European countries may overwhelmingly represent white skin types that have low melanin content as opposed to dark skin tones with higher melanin, leading to incorrect results in a country like India.</p>
Biases in historical data	<p>Even if protected attributes like gender or race are removed, data could have bias due to historical reasons.</p> <p>For example, a <u>hiring algorithm</u> by Amazon favoured applicants based on words like “executed” or “captured” that were mostly used by men in their resumes. Learning from this, the algorithm started preferring men over women and even dismissed resumes with the word ‘woman/women’ in them. Amazon eventually stopped using the algorithm.</p>
Use of poor-quality data	Poor predictions may also be the <u>result of</u> low-quality, outdated, incomplete or incorrect data at different stages of data processing.



Checklist for Developers at Different Points of Intervention

Pre-Processing

Are you able to identify the source/sources of bias at the stage of data collection?

Did you check for diversity in data collection before it was used as training data to mitigate bias?

Did you analyse the data for historical biases?

In-processing

Have you assessed the possibility of AI correlating protected attributes and bias arising as a result?

Do you have an overall strategy (technical and operational) to trace and address bias?

Do you have technical tools to identify potential sources of bias and introduce de-biasing techniques? Please see Appendix for a list of technical tools that developers may consider

Have you identified instances where human intervention would be preferable over automated decision making?

Post-processing

Have you identified cases where human intervention will be preferred over automated decision making?

Do you have internal and/or third-party audits to improve data collection processes?



Good Practices

- Developers may use methods like relabelling, reweighing, data transformation to eliminate correlation between protected attributes or appropriate technical methods to ensure that the data samples are fair and representative enough to give accurate results.
- Ensure that training samples are diverse to avoid racial, gender, ethnic, and age discrimination or redact protected attributes when collecting data.
- Developers can ensure multiple and diversified human annotations per sample and that those annotators come from diverse backgrounds.
- Developers could measure the outcomes for different ethnic/racial and/or gender groups to see if there is unfair treatment.
- Developers could actively collect more training data from historically under-represented groups that may be at risk of bias and apply de-biasing techniques to penalize errors.
- Regular audits of the production models for accuracy and fairness, and regularly update the models using newly available data.
- Apply bias mitigation techniques such as [adversarial debiasing](#), [Fairness constraints](#), [prejudice remover](#) [regularizer in process](#).



At a Glance

Bias becomes an evident concern only once the AI system starts producing outputs. However, measures to mitigate bias can be adopted at multiple stages of processing. The developer's prerogative is to anticipate potentially biased outcomes before they arise. The source of bias could be the data used during training/processing, the attributes used by the AI system or the functions for which the AI is designed for. The developer should be aware of potential biases in each of these aspects. Some measures that developers can practice in the interest of mitigating bias include:

Regular audit of data and labels to ensure diversity in data and attributes used.

Awareness exercises on historical and contextual biases and discussion on technical protocols for bias mitigation.

Technical methods to ensure fairness such as reweighting, relabelling and data transformation to eliminate any unanticipated correlation.

Outcome measurement and comparison of results across identities to ensure equally accurate outcomes.

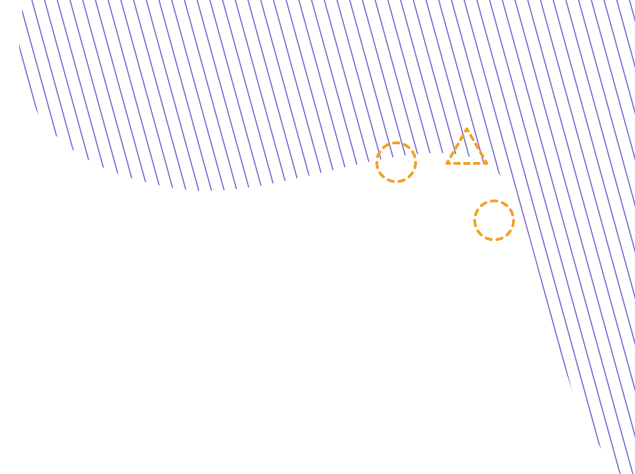
Instances of challenges developers may face

Multilingual models for training data are significant challenges that developers face and is also a source of bias percolation. In most cases, English is the language of preference and the one that is easily referenced. This significantly impairs the ability to trace bias in inputs from other languages.

In India, data collection is mostly done in urban areas and from segments of the society or specific regions. These do not include other dialects from rural parts. The data overrepresents dominant communities. As a result, although the product is meant for every Hindi user, the product will not be a good fit as it lacks the due diligence of representing data adequately.

In case of facial recognition, insufficient number of images or false correlations drawn by the AI enabled system could lead to bias in classifying images. There are numerous instances of facial recognition being faulty in settings involving minorities and instances where the software was applied to subjects that were not adequately represented in training data.

The COMPAS software and the [Allegheny Family Screening Tool](#) in the United States were found to be biased against African-American and biracial families. The biases of the model stem from a public dataset that reflects broader societal inequities. By relying on publicly available data, those who had no alternative to publicly funded social services, predominantly low-income families, were over-represented in the data.



Further Questions

What kind of algorithms are more likely to amplify biases?

Any algorithm can develop a bias. So far, the most common example/instances of bias that we can identify are in (i) Online recruitment tools: The case of Amazon's hiring tool is mentioned above; (ii) Word associations: Princeton University [research](#) found that an algorithm was associating the words 'woman' and 'girl' with arts and humanities and men with science and math; (iii) Bias in online advertisements: [Research](#) by the Federal Trade Commission in the United States found that online search queries for African-American names were most likely to produce arrest records, while for white males, the search results were different; (iv) Facial recognition bias: One such case has already been discussed in the chapter. (v) Criminal justice algorithms: The example of COMPAS software describes such biases.

What is the impact of biased algorithms?

Bias directly results in unequal allocation of opportunities, resources, or information. In a world that is adopting AI in many important public functions like ration distribution, and determining eligibility for welfare schemes, and other crucial functions like granting loans, bias towards a community that is already marginalised can have disastrous consequences. It could also have a very direct impact on individuals and their well-being.



IV. Fairness

Meaning

Fairness is one of the most relevant and the most difficult topics in AI ethics. It is a normative concept and has numerous definitions. Arvind Narayanan at the Association for Computing Machinery Fairness, Accountability and Transparency (ACM FAccT) Conference in 2018 presented some of these definitions for fairness. The work understands the social context of fairness in a technical manner and seeks to balance accuracy with fair outcomes. Broadly, these definitions say that developer must actively include checks and balances during algorithmic design to ensure that there is no individual or group discrimination in outcomes of the AI process.

For ensuring fairness, developers need to follow ethical values and assess whether AI systems impact people adversely. For example, AI may unfairly allocate work opportunities where women candidates are rejected or AI facial recognition systems may not provide the same quality of service to users with a particular skin tone. These are biased outcomes and the role of fairness as an ethical imperative is to ensure that these outcomes are not overlooked for the sake of accuracy. Fairness addresses the ethical dilemma of choosing between what is a fair outcome and what is an optimal outcome.

Fairness helps developers understand what values cannot be compromised in algorithms. Academics refer to Fairness along with Accountability and Transparency as the FAT framework. Industry has also engaged with fairness frameworks. In 2018, Accenture published a fairness toolkit to assist companies work towards fairer outcomes when deploying AI systems. An established framework/toolkit provides developers an advantage as it is a ready-reference for gauging fairness. There are also case-studies by companies such as Microsoft and Spotify that demonstrate implementational challenges with fairness toolkits.

Scenario:

A loan applicant comes to know that her bank has denied a loan application to start a new business. A male friend of hers with the same credit rating and similar economic background is granted the loan. The bank tells her that their AI enabled system considered her application and rejected her loan. On further investigation, it is revealed that the AI enabled system denies loans to all unmarried women. The start-up that supplied the AI technology is sure that this is the optimal outcome, and the bank should accept the decision suggested by the AI.

How do you address fairness in this scenario?

The technology provider in this case is sure that the outcome produced by the AI is accurate. Despite the scope for bias in training data, the case may be that denying this loan ensures that there is no loan default. However, it is logically incorrect to say that an outcome is accurate when it is ignoring a reality that unmarried women have not been granted loans before by the bank. Here, there is a need to reconcile the accuracy rate of the AI enabled system with social realities. The thumb rule is that an AI-enabled system should not produce disproportionately accurate or inaccurate results for one group, in comparison to the other. False positives and false negatives³ must not vary across social, ethnic, gender, caste, class groups.

³ False positive is an error when something does not exist, but the result shows it does and a false negative is when something exists, but the result does not show it. For example, in Covid testing a false positive is when someone shows positive when they do not have the virus and a false negative is when the test shows up as negative, but the person is actually infected. It is an error quotient.



Checklist for Developers at Different Points of Intervention

Pre-Processing

Have you anticipated the possibility of the algorithm treating similar candidates differently on the basis of attributes such as race, caste, gender or disability status?

Are there sufficient checks to ensure that the machine does not base its outputs on protected attributes?

Post-processing

Does the algorithm provide results that are equally and similarly accurate across demographic groups?

Have you checked outcomes to understand if the AI is producing equal true and false outcomes? Is there any scope for disproportionate outcomes?

In-processing

Are there protocols or sensitisation initiatives to reconcile historical biases and build in weights to equalise potential biases?

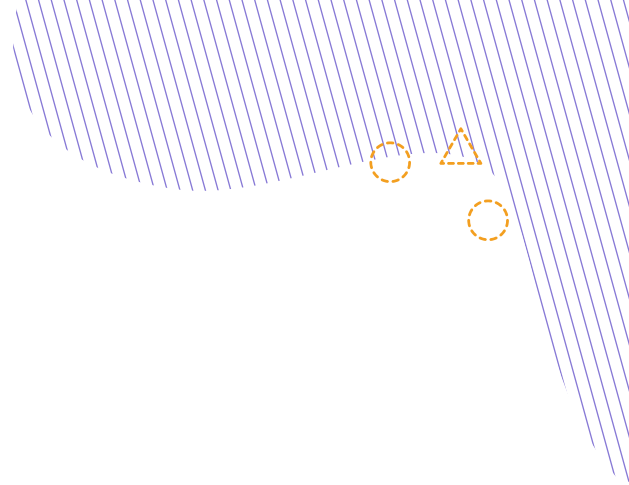
Have you conducted due diligence to trace potential fairness harms that may arise directly or indirectly?

Do you have technical tools to diagnose fairness harms and address them?



Good Practices

- Identify communities and groups who are vulnerable to fairness-related harms. Look at people who: (i) will use the AI system; and (ii) are directly or indirectly affected by the system (by choice or not). Next, the effect of AI decision-making should be tested against anti-discrimination laws and safeguards.
- A context specific approach to fairness where possibility of discrimination is identified based on context and historical biases. At the first stage, this involves looking at the purpose and context for which AI is deployed. The developer can then predict if the AI system will produce unfair outcomes when assessing similar candidates because of their gender, race, caste or disability.
- It is also important to acknowledge that individuals may belong to overlapping groups i.e., different combinations of gender, race, caste or disability. They may be especially vulnerable to fairness related harms and different kinds of harms. Groups should be considered both separately and collectively to expand the understanding of harms.
- Refer to fairness existing toolkits such as [this](#).



At a Glance

Fairness is a complicated subject with multiple definitions. Fairness cannot be separated from the context in which AI is applied so it is difficult to have overarching principles to determine fairness across a wide range of AI applications. The starting point of fairness is to ensure non-discriminatory outcomes. In short, the AI system should provide equally and similarly accurate results across demographic groups. To ensure fairness, a combination of the following may be considered

A map of anti-discrimination safeguards in law and functions of the system that should be mindful of these measures.

A policy on fairness in AI outcomes and a mandate to have the same range of accuracy across demographic groups.

Fairness checks at the envisioning stage, pre-mortem stage and product-greenlighting stage, or at pre-processing, in-processing and post-processing stages.

To incorporate fairness, developers must familiarise themselves with legal concepts mentioned below.

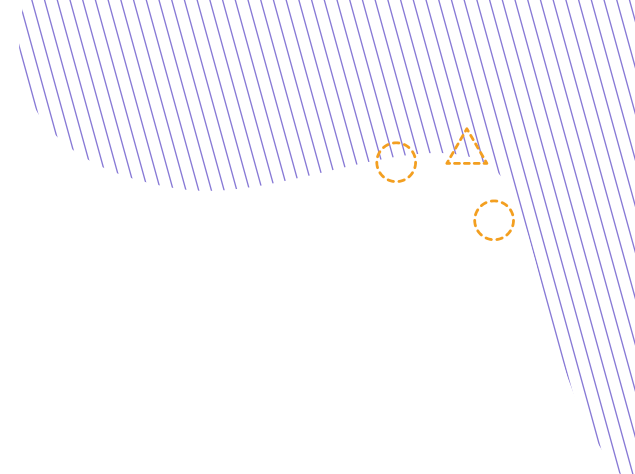
Transparency - it helps analyse how data processing will impact its users. Refer to the chapter on Accountability and Transparency.

Bias mitigation - it reduces the extent of harm people may experience by interacting with AI services and products. Developers must refer to chapters on data quality, rights of data principals and safeguarding sensitive personal data.

Protection of personal data and familiarity with rights of data principals - guides developers to secure data and avoid processes that are detrimental, discriminatory or misleading. For details, refer to Rights of Data Principals and Personal Data.

Assessing whether you are processing information fairly depends partly on how you obtain it. If anyone is deceived or misled when the personal data is obtained, then this is unlikely to be fair.

To assess whether or not you are processing personal data fairly, you must consider more generally how it affects the interests of the people concerned – as a group and individually. If you have obtained and used the information fairly in relation to most of the people it relates to but unfairly in relation to one individual, there will still be a breach of this principle.



Instances of challenges developers may face

Developers may use post-processing algorithms to reduce unfairness. Some challenges here are that post-processing algorithms may need access to sensitive data for calculating different thresholds for different groups, in case of loan grants for example, to predict their loan repayment ability. However, the use of sensitive data may be prohibited under law and developers may not be able to use the same.

Without a framework or a toolkit for reference, it is difficult to anticipate future discrimination when the AI system evaluates two similar candidates. This could lead to discriminatory outcomes that undermine fairness.

If a voice recognition system is mainly trained on male voices, the system may not perform accurately when used by women. The quality of the data used leads to unequal performance of the services for different groups in the population.

Further Questions

What does a fair approach mean?

A fair approach would recognise or account for structural and systemic inequalities, constitutional guarantees of affirmative action and the responsibility to correct past unfair practices. As stated earlier, the fundamental idea of fairness is to ensure that fairness is not overlooked for objective accuracy. Fairness is an evolving concept, and it is paramount to build in checks and balances for fairness, given the increasing role of AI in society.

What differentiates fairness and bias as two distinct ethical values?

A bias is like a bug in a code. It is an unwarranted association made by the AI-enabled system, which may be deliberate in the design or an unintended association. Fairness is ensuring that the outcomes do not differentiate between two groups or even two individuals. Bias may be one of the reasons for unfair outcomes, but all unfair outcomes are not the result of a bias. A bias is always unfair. Bias is a design issue while fairness is a design plus outcomes issue, and both need separate attention and separate approaches.

V. Security

Meaning

Safety and security while using AI systems are important to preserve public trust in AI. As technology advances, so do the threats to security. Response to such threats should also grow accordingly. The need for secure networks is not overstated, and more than 50 per cent organizations in India suffered from a cybersecurity breach in 2020. Many start-ups including bigger organisations like BigBasket and Flipkart, highlight that start-ups should be especially mindful of security as this could be a factor in their valuation, trust factor and general robustness.

Security immunizes the user or beneficiary of an AI system from digital risks. This means that they should be protected from unreasonable security risks, digital and physical, during the foreseeable use of the technology. A developer should keep in mind the following fundamentals in AI security: (i) be

aware of attempts to manipulate outcomes with inputs that appear legitimate; (ii) design the AI-enabled system and the development process to anticipate and mitigate security risks; (iii) incorporate general cybersecurity principles in AI design; (iv) limit those who can control or make changes to AI-enabled systems and create a register of log-ins.

Developers can build secure models if they are able to pre-empt security concerns and address them effectively. The best practice here is to conduct a risk management audit to anticipate foreseeable misuse, i.e., risks associated with the use of an AI system in other contexts. The Organisation for Economic Co-operation and Development (OECD) suggests standards for digital security risk management and risk-based due diligence under the MNE Guidelines and OECD Due

Diligence Guidelines for Responsible Business Conduct. For start-ups, ensuring security means adoption of proper security standards to ensure high safety and reliability while also limiting the risk exposure of developers. This is a difficult bargain for an AI technology start-up with limited staff and limited resources.

The role of AI in security has both advantages and disadvantages. On one hand, AI is a very useful tool in real time discovery of phishing and spam in password protection and user authentication, and tracing fake news and deepfakes. On the other hand, AI also enables efficient, targeted attacks on a scale previously unimaginable. The role of security in AI ethics is to ensure robust networks that prevent unauthorised access that could compromise the system and move AI's decision-making away from the control of developers.

Scenario:

A loan applicant applies for a loan to start her business and the bank approves the loan. In her loan application, she said that the purpose of the loan was to start a fertilizer business. From the next day, she starts receiving calls and email advertisements for purchasing wholesale fertilizers, rubber gloves etc. When she approaches the bank, the bank tells her that they do not share personal information with third parties. The AI technology provider also says that they do not harvest and sell data. On further diagnosis, it is revealed that someone has breached the AI enabled system and stolen data including sensitive data.

How does security help?

Security means inculcating practices and protocols that ensure that the network remains robust and free from outside interference. When AI is deployed in settings that involve significant public interest, ensuring secure networks also have geo-political angles to it. Security protects the user/beneficiary and also protects the organisation from financial risk. There have also been security attacks where the attackers seek money in exchange for returning control of an AI system. Security, as an ethical imperative, mitigates the damage caused by such instances to some extent.



Checklist for Developers at Different Points of Intervention

Pre-Processing

Have you identified scenarios where safety and reliability could be compromised, both for the users and beyond?

Have you classified anticipated threats according to the level of risk and prepared contingency plans to mitigate this risk?

Have you defined what a safe and reliable network means (through standards and metrics), and is this definition consistent throughout the full lifecycle of the AI-enabled system?

In-processing

Have you created human oversight and control measures to preserve the safety and reliability risks of the AI System, considering the degree of self-learning and autonomous features of the AI System?

Do you have procedures in place to ensure the explainability of the decision-making process during operation?

Have you developed a process to continuously measure and assess safety and reliability risks in accordance with the risk metrics and risk levels defined in advance for each specific use case?

Post-processing

Have you assessed the AI-enabled system to determine whether it is also safe for, and can be reliably used by users with special needs or disabilities or those at risk of exclusion?

Have you facilitated testability and auditability?

Have you accommodated testing procedures to also cover scenarios that are unlikely to occur but are nonetheless possible?

Is there a mechanism in place for designers, developers, users, stakeholders and third parties to flag/report vulnerabilities and other issues related to the safety and reliability of the AI enabled System? Is this system subject to third-party review?

Do you have a protocol to document results of risk assessment, risk management and risk control procedures?



Good Practices

- Ensure that vendors follow security protocols after contracts for services are signed. Regularly check if technology partners are keeping their promises to protect personal information.
- Follow up with technology contractors to ensure they follow security protocols. If a vendor is supposed to delete extraneous information, ask for proof.
- Participate in international standard setting processes such as International Organisation for Standardisation (ISO) and the International

Electrotechnical Commission (IEC) standards for the development and deployment of safe and reliable AI systems.

- ISO/IEC 27001 series provides guidance on protection of privacy, irrespective of the type and size of organisations. The latest standard [ISO 27003](#) adds an extra layer of security to the ISO 27001 standard and helps organisations design, build, implement and continually improve their personal information management systems.



At a Glance

AI systems should be robust, secure and safe throughout their entire lifecycle. This requires developers to look at the conditions of use presently, foreseeable use or deployment in unintended operations to minimise safety risk. Mainly, they should observe the following:

Ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle.

Analyse the AI-enabled system's outcomes and responses to ensure contextual and state of the art AI.

Based on their roles, the context, and their ability to act, developers should apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis.

Be mindful of risks related to AI-enabled systems, including privacy, digital security, safety and bias.

Instances of challenges developers may face

In April 2018, OpenAI released the OpenAI Charter. Among other things, it states that their objective is to publish research that benefits society. According to this Charter, safety and security is a factor that should determine whether you should publish AI research or not. The decision to release GPT-2 in a staggered manner stemmed from this Charter and the concern that the model could be used maliciously, primarily for the generation of scalable, customizable, synthetic media (deepfakes and augmented videos). For example, OpenAI noted that its model could be used to generate misleading news articles, impersonate others online, automate the production of abusive content online, and automate phishing content.

In November 2019, the cybersecurity company FireEye published a blog post revealing they were using GPT-2 to detect social media posts as part of information operations. The company's researchers fine-tuned the GPT-2 model and taught it to create tweets that resembled the source. While researchers were able to detect malicious activity such as bot tweets, they also realised that the model could also lower the entry barrier to entry for bad actors if rolled out without more security protocols.



Further Questions

Is following the ISO standard enough to protect the AI-enabled system?

An ISO standard is a minimum technical standard. It is also important to have processes and protocols to systematically review the robustness of the network. The best standard of protection is to have an external auditor or cybersecurity provider secure the system. It is also important to ensure that security protocols, procedures and auditing is followed from the design stage to the implementation station to ensure maximum security and safety. See this [article](#) for a detailed explanation on why it is beneficial for start-ups to go beyond the basic compliance standard.

What is the difference between privacy and security?

Privacy is the obligation to not misuse data, and security is the obligation to protect the entire AI-enabled system including data. A security breach could result in data being stolen which then becomes a violation of privacy. Developers need to understand that these are two distinct obligations and address them separately.



VI. Privacy

Meaning

Privacy can be understood simply as the right to be left alone. It is individual autonomy over sharing information about oneself, with others and the public at large. In India, the right to privacy is a fundamental right under Article 21 of the Indian Constitution. The Supreme Court recognised the right to privacy as a part of the fundamental right to life and liberty in 2017, in Justice K.S. Puttusawmy and Ors. v. Union of India. Jurists conceptualise privacy in

numerous ways. Among others, it means (i) bodily inviolability and integrity and intimacy of personal identity including marital privacy; (ii) the right to be left alone⁴; and (iii) a zero relationship between two or more persons in the sense that there is no interaction or communication between them, if they so choose⁵.

Privacy interacts with AI when large amounts of customer and vendor data are fed into algorithms to generate insights

without the knowledge of data principals. Privacy is also violated at the stage of data collection, if appropriate consent is not obtained before collecting and aggregating data for model training. However, it is important to note that AI models can learn only with data. This is the basis for the ethical and social dilemma for developers - balancing privacy against technological breakthroughs and improved services.

Scenario:

A loan applicant applies for a loan to start her business and the bank approves the loan. In her loan application, she said that the purpose of the loan was to start a fertilizer business. From the next day, she starts receiving calls and email advertisements for purchasing wholesale fertilizers, rubber gloves etc. When she approaches the bank, the bank tells her that they do not share personal information with third parties. On further investigation, it is revealed that the AI enabled system that determines whether or not loans should be granted is selling personal information to third parties.

How does privacy help?

As defined above, privacy is the right to be left alone. In this scenario, this means that a loan applicant should be left alone and not disturbed with advertisements that she has not subscribed to. Privacy law is evolving at a rapid pace and some of the principles that have emerged are (i) informed consent should be taken from data principals before sharing data; (ii) the data should only be used, processed for the purposes that the subject has given consent for; and (iii) data should be deleted after the period of time for which it will be used. Mechanisms to ensure this is explained in this section and in detail in the legal section (Part II of this Handbook).

⁴ Cooley, Thomas M.A Treatise on the Law of Torts, 1888, p.29 (2nd ed.)

⁵ Shils Edward, Privacy, Its Constitution and Vicissitudes, 31 Law & Contempt Problems, 1966, p.281



Checklist for Developers at Different Points of Intervention

Pre-Processing

Have you considered deploying privacy specific technological measures in the interest of personal data protection? Please see the Appendix for an illustrative list of technological measures you could consider.

Are you able to trace the source for the data being used in the AI system? Is there adequate documentation of informed consent to collect and use personal data for the purpose of developing AI?

Is sensitive data collected? If so, have you adopted higher standards for protection of this kind of data?

Does the training data include data of children or other vulnerable groups? Do you maintain a higher standard of protection in these cases?

Is the amount of personal data in the training data limited to what is relevant and necessary for the purpose?

In-processing

Have you considered all options for the use of personal information (e.g. anonymisation or synthetic data) and chosen the least invasive method?

Have you considered mechanisms/ techniques to prevent re-identification from anonymised data?

Have you been informed of the legal requirements for processing personal information? What measures does your organisation adopt to ensure compliance?

Post-processing

Are there procedures for reviewing data retention and deleting data used by the AI System after it has served its purpose? Are there oversight/review mechanisms in place? Is there scope for external/third-party review?

Beyond the data principal privacy, have you evaluated the potential of data of an identified group being at risk?

Is there a mechanism in place to manage consent provided by data principals? Is this mechanism accessible to principals? Is there a method to periodically review consent?



Good Practices

- Before processing personal information, anticipate the impact of the process on data protection and the likelihood of a risk to the rights and freedoms of natural persons. In case of new technology such as Artificial Intelligence, this is paramount and the nature of the processing, its scope and purpose, and the context in which the technology will be implemented needs close evaluation, even third-party or public scrutiny.
- Take best efforts to adopt a multidisciplinary approach. AI is more than just technology. It is important to put together multi-disciplinary approaches that can consider various perspectives on the consequences of technology on society.
- Limit the amount of personal data in the training data to what is relevant and necessary for the purpose.
- Ensure and document that the AI system you are developing meets the requirements for privacy by design.
- Document how the data protection requirements are met. Documentation is one of the requirements of the regulations and will be requested by customers or users.
- Ensure good systems for protecting the rights of data principals, such as the right to information, to access and deletion. If consent is the legal basis of processing, the system must also include functionality enabling consent to be given, and to be withdrawn.
- Ensure that vendors follow security protocols after contracts for services are signed. Regularly check if technology partners are keeping their promises to protect personal information.
- Follow up with technology contractors to ensure they follow security protocols. If a vendor is supposed to delete extraneous information, ask for a proof of deletions.
- Protocols for user-system interaction, and novel methods to segregate data through specific minimisation requirements are some techniques that can be adopted without compromising quality. Interdisciplinary associations with legal and other social sciences are also some useful techniques.



At a Glance

Privacy or data protection is critical to any AI business from a legal and an ethical perspective. The legal mechanism for regulating data protection is in the final stages of deliberation in many countries including India. Some data such as financial data, health data, genetic data or children's data is accorded higher protection than other personal data. Developers should be mindful of this. In the interest of privacy, developers should consider the following

Methods for reducing the need for training data and minimising data use.

Methods that uphold data protection without reducing the basic dataset.

Measures to track consent for data that is used and redressal or review mechanisms to provide data autonomy to data principals.

Adhere to data protection regulations and best implementation practices.

Instances of challenges developers may face

The Norwegian Tax Administration (NTA) developed a predictive tool to assist in selecting individuals for random checks for tax evasion and filing errors. They tested 500 different variables including the taxpayers' demography, life history, and other details in their tax returns. Finally, only 30 variables were built into the final model. This provides a good example of how it is not always necessary to use all the available data in order to achieve the desired purpose. However, this requires time and effort to streamline the product.



Further Questions

Why is it important to understand different classification of data?

Data that is fed into an AI system can reveal much more information than what is required for the purpose for which the AI system was made. Sensitive data that reveals a person's gender or race may increase the possibility of bias and distort the product's outputs. This may pose significant risks to the AI system as it will not be considered reliable for use. Therefore, the developers must learn to identify what data they are using so that they can develop appropriate methods to prevent its misuse. The law classifies data such as health records, biometric data, caste, religious or political beliefs, bank details, etc. as sensitive personal data.

Why are there sector-specific data protection guidelines?

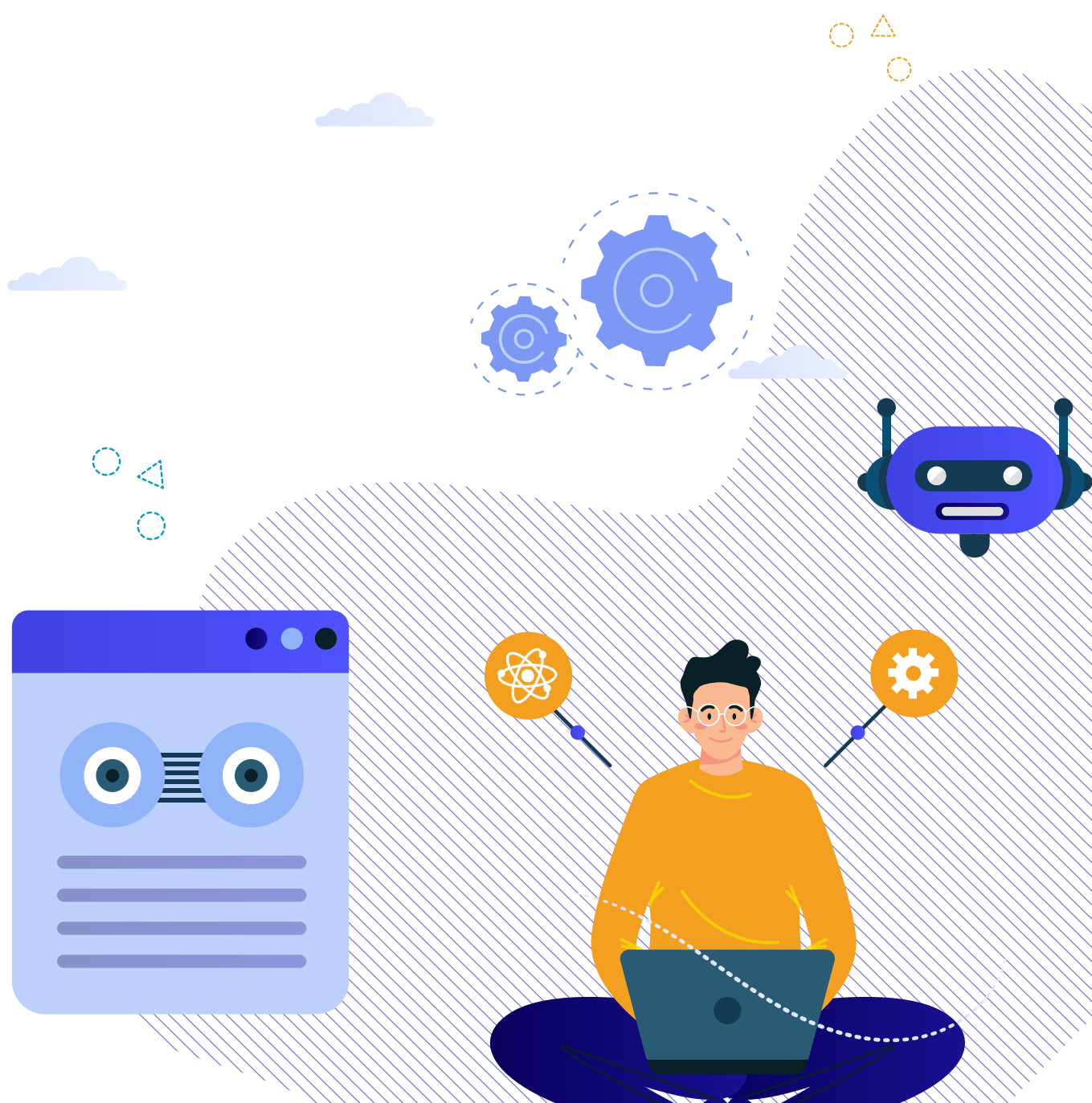
Sector-specific data protection guidelines emerge as a response to large scale use of certain data in an unregulated environment which could potentially harm data principals. The two biggest examples of such guidelines are for fintech companies and in the health sector. Financial data and health data are accorded a higher degree of protection because lax protection could have severe ramifications on individuals and because of the scale of data use in emerging technology sectors such as AI/ML.

Ethics in AI Literature and Policy Documents

- [NITI Aayog National Strategy for AI](#)
- [OECD Principles on AI - Human-centered values and fairness](#)
- [High-Level Expert Group on Artificial Intelligence: Ethical Guidelines for Trustworthy AI](#)
- [Singapore Model AI Governance Framework \(Second Edition\)](#)
- [United Kingdom's Ethics, Transparency and Accountability Framework for Automated Decision-Making](#)
- [United Kingdom's Data Ethics Framework](#)
- [Explainable AI: The Basics Policy briefing, The Royal Society](#)
- [The 8 Principles of Responsible AI - ITechLaw](#)
- [Asilomar Principles on Beneficial AI](#)
- [Montreal Declaration for responsible development of AI](#)
- [The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems](#)
- [Tenets - Partnership on AI](#)
- [Five Overarching Principles for an AI Code - UK House of Lords](#)
- [Statement on artificial intelligence, robotics and 'autonomous' systems - European Parliament](#)
- [AI Fairness Toolkit](#)
- [The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems](#)
- [FAT/ML: Principles for Accountable Algorithms and a Social Impact Statement for Algorithms](#)
- [Algorithmic Accountability Bill of 2019 in USA](#)
- [Software and Information Industry Association's Ethical Principles for Artificial Intelligence and Data Analytics](#)
- [Algorithmic Impact Assessment Framework - AINow Institute](#)

Section II

Data Protection



Introduction

Artificial Intelligence is data intensive and due to this reason, data protection laws assume great significance for AI professionals. The recognition of the right to privacy as a fundamental right by the Supreme Court of India has resulted in greater emphasis on ensuring that data collection, storage, sharing and processing practices respect user privacy. Presently, India is at the cusp of adopting a new data protection framework. This includes the Personal Data Protection Bill, 2019 (PDP Bill), which is pending before the Parliament, and a proposed non-personal data framework, which is at an

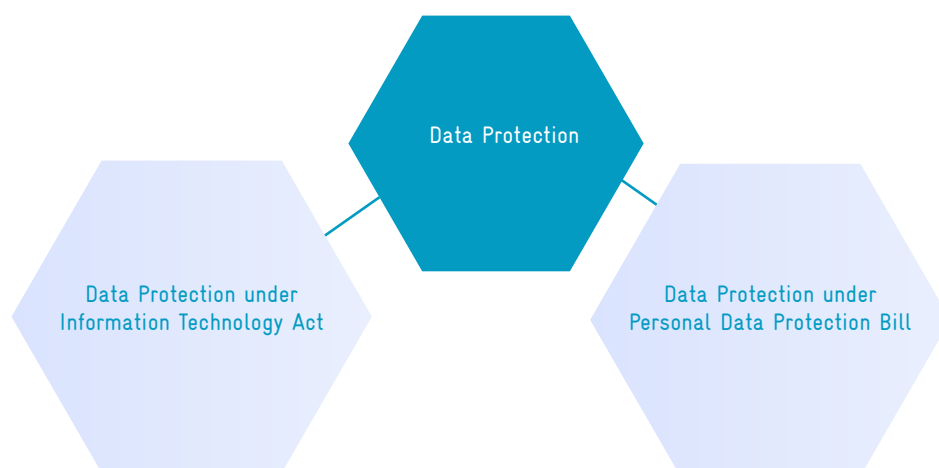
early stage of its lifecycle. The proposed laws will replace the existing framework consisting of the Information Technology Act, 2000 (IT Act) and the Information Technology (Reasonable security practices and procedures and sensitive personal data or information) Rules, 2011. This will lead to new compliance requirements, more stringent implementation provisions and greater penalties.

There is a direct correlation between data protection framework and the quality of data collected. Under a weak data protection framework, an individual concerned about his privacy will have incentive

to furnish incorrect information about himself. This will lead to poor data quality, which will then feed into poor data analytics and ultimately impact businesses too. Similarly, a stringent data protection framework can incentivise individuals to furnish more accurate personal information.

This chapter will highlight the broad principles enshrined in the proposed framework, with the objective of getting AI professionals ahead of the curve. However, before describing the proposed framework, it is important to have a look at our existing data protection framework.

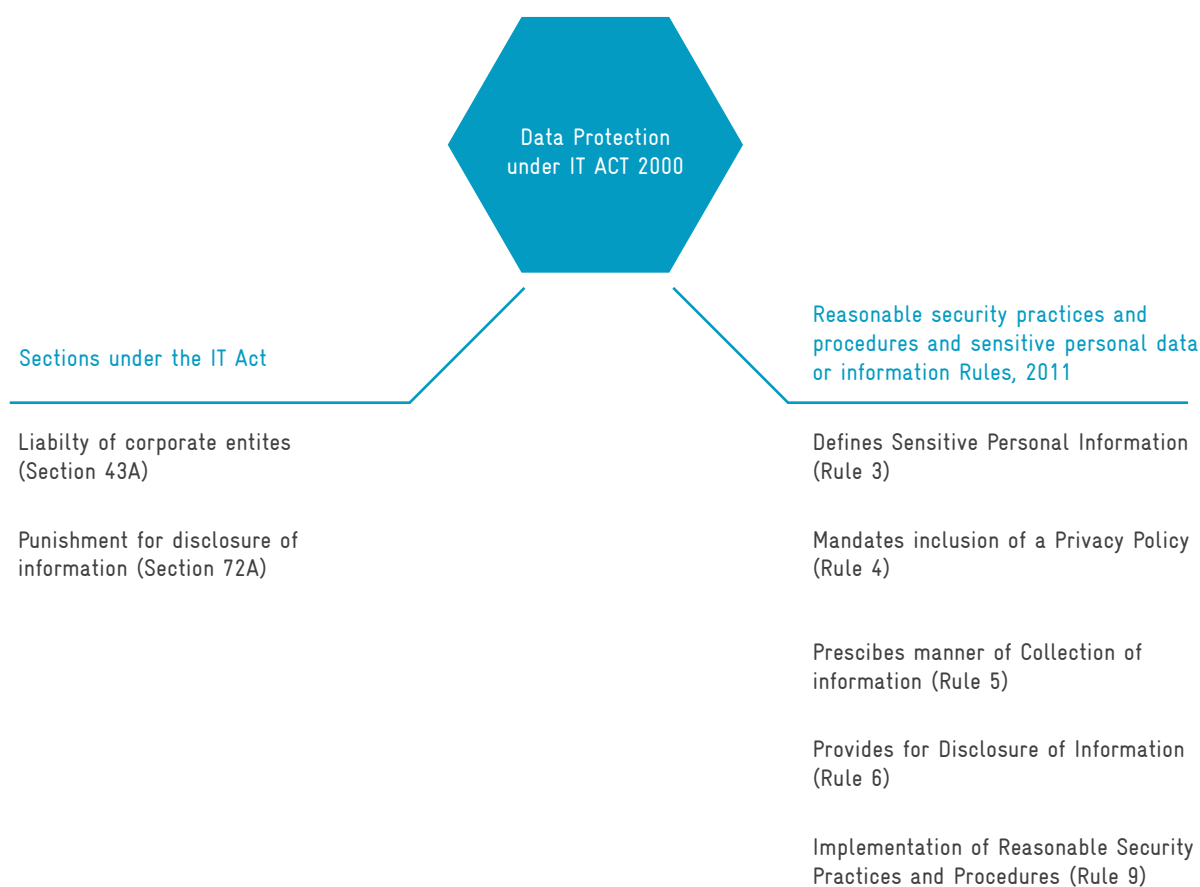
Figure 3 : Existing data protection framework



Data Protection under Information Technology Act, 2000

Until the PDP Bill is passed by the Parliament, India's data protection framework will continue to be governed by the Information Technology Act, 2000 and the Information Technology (Reasonable security practices and procedures and sensitive personal data or information) Rules, 2011.

Figure 4 : Provisions relating to personal data protection under the IT Act



Sections under IT Act

There are two broad provisions under the IT Act that relate to data protection.

I. Liability of corporate entities

- *Onus is on corporations for protecting sensitive personal data:* This is given under Section 43A of the IT Act. The corporate entities will be held liable for negligence in maintaining and implementing reasonable security practices while processing sensitive personal data, if it leads to wrongful loss (loss of property to which the person losing is legally entitled) or wrongful gain (gain of property to which the person gaining is not entitled) to any person. In such cases, the corporate will have to pay compensatory damages to the aggrieved person.
- Punishment for disclosure of personal information if there is a breach of contract. This is given under Section 72A of the IT Act. The clause punishes a person with imprisonment in case he discloses personal information in breach of contract or without consent

of the person to whom it belongs. It states that any person who is providing a service under a contract, discloses personal information obtained during the course of providing such services, without the person's consent or in breach of terms of the contract, will be punished with either imprisonment for a term not exceeding three years or with a fine not exceeding up to five lakh rupees, or both.

II. Reasonable security practices and procedures and sensitive personal data or Information Rules, 2011

Every entity that owns or processes data must implement reasonable security practices. These practices are given under rules known as the Information Technology (Reasonable security practices and procedures and sensitive personal data or information) Rules, 2011 [SPDI Rules]. The Rules provide details about how sensitive data must be handled.

i. Sensitive Personal Information

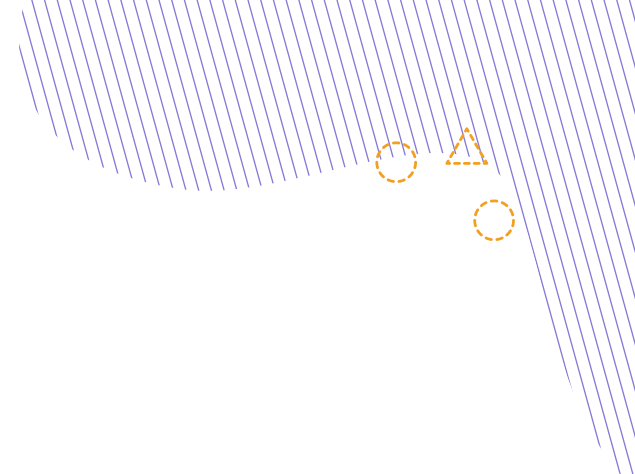
The definition of sensitive personal information is given

under Rule 3. It says that sensitive personal data or information of a person means personal information that consists of information relating to:

1. password;
2. financial information such as bank account or credit card or debit card or other payment instrument details;
3. physical, physiological, and mental health condition;
4. sexual orientation;
5. medical records and history;
6. biometric information;
7. any detail relating to the above clauses as provided to body corporate for providing service; and
8. any of the information received under any of the above clauses by body corporate for processing, stored or processed under lawful contract or otherwise.

ii. Privacy Policy

This is given under Rule 4. According to this Rule, all entities that are seeking sensitive personal data have a duty to a privacy policy. This policy must be made easily accessible for people who are providing their information. The policy must be published on the website of the entity collecting such data and give details on the



- Type of information that is being collected
- Purpose of collection
- Security practices in place to maintain the confidentiality of such information.

iii. Collection of Information

Rule 5 provides the guidelines that need to be followed by a body corporate⁶, while they collect information. The Rule imposes the following duties on the Body Corporate:

- Obtain consent from the person(s) providing information.
- Information shall only be collected for a lawful purpose and only if considered necessary for the purpose.
- It will be used only for the purpose for which it is collected and shall not be retained for a period longer than which it is required;
- Ensure that the person providing information is aware of the fact that the information is being collected. He/she must also be aware of its purposes & recipients, name and

addresses of the agencies retaining and collecting the information;

- Offer the person an opportunity to review the information provided and make corrections, if required;
- Before collection of the information, provide an option to the person providing information to not provide the information sought;
- The provider of information shall also have an option to withdraw their consent given earlier;
- Maintain the security of the information provided; and
- Designate a Grievance Officer whose name and contact details should be on the website and who shall be responsible to expedite and address grievances of information providers expeditiously.

iv. Disclosure of Information

Under Rule 6, disclosure of sensitive personal information to any third party will require prior permission from the

person providing the information. However, no prior permission is required if a request for such information is made by government agencies authorised under law or any other third party by an order under law.

v. Reasonable Security Practices and Procedures

Rule 8 provides the reasonable security processes and procedures that may be implemented by the collecting entity. International Standards (IS / ISO / IEC 27001) is one such standard which can be implemented by a body corporate to maintain data security. An audit of reasonable security practices and procedures shall be carried out by an auditor at least once a year or as and when the body corporate or a person on its behalf undertakes significant upgradation of its process and computer resource.

⁶ Body Corporate includes entities like a company, co-operative society, registered societies and firms etc.

Data Protection under Personal Data Protection Bill, 2019

Glossary

Personal Data

Any information using which a person can be identified either directly or indirectly is personal data. Examples of personal data include, name, age, gender, medical records, bank account details, etc., of a person.

Data Principal

Person to whom the personal data relates. For example, if you collect the age, gender and blood group of a person named 'X', X will be the data principal.

Data Fiduciary

A person or entity who decides the purpose and means of processing personal data.

Data Processor

A person or organisation that processes personal data on behalf of a data fiduciary is a data processor.

Processing

Processing refers to operations or sets of operations performed on personal data. It includes, among other activities, the collection, storage, organisation, alteration, retrieval, usage, distribution, and erasure of personal data.

Significant Data Fiduciary

A data fiduciary may be classified as a significant data fiduciary based on a number of factors including the volume of data processed, sensitivity of data processed, risk of harm, etc.

Sensitive Personal Data

Personal data which may reveal, be related to, or constitute financial data, health data, biometric data, genetic data, sexual orientation, sex life, caste or tribe etc.

Principles of Personal Data Protection

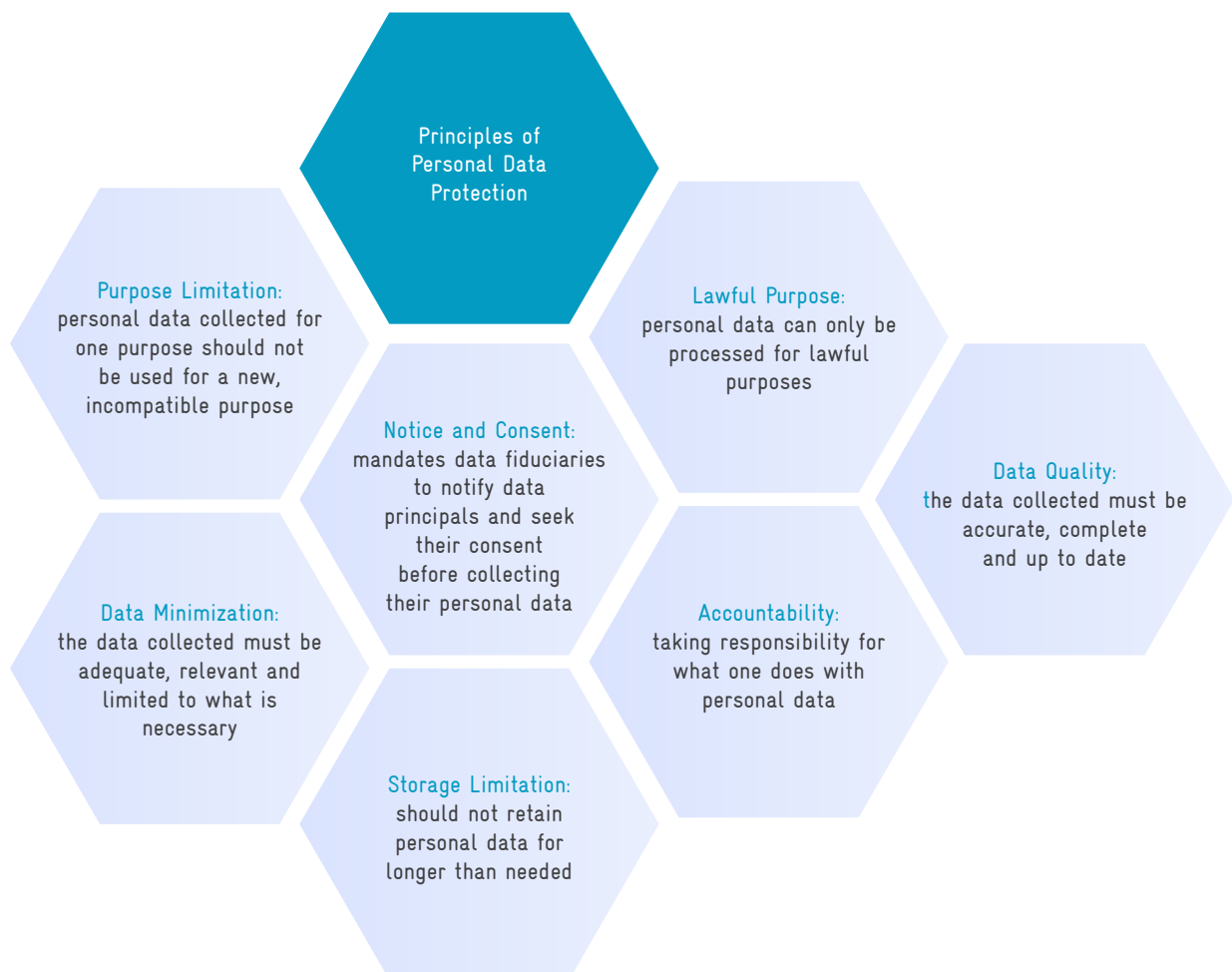
The proposed data protection framework has not yet been formalised into law. Hence, the precise provisions which the said laws will entail cannot be ascertained. However, the principles on which the new framework is based are certain. These principles embody the

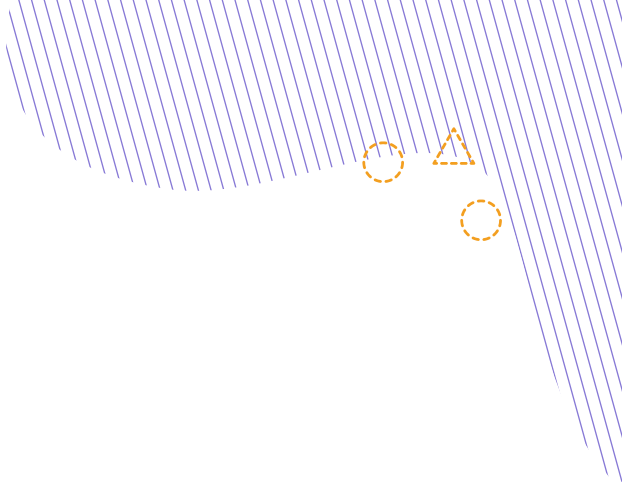
spirit of India's approach to protecting personal information and upholding privacy. Thus, understanding and complying with these principles is at the heart of good data protection practices. Failure to comply with these principles will leave you open

to penal consequences involving substantial fines and imprisonment.

Following are the key principles enshrined in the proposed framework.

Figure 5 : Principles of personal data protection





Lawful purpose

The law requires that processing of personal data can only be done for lawful purposes. While the term is not defined, in common parlance, a purpose is lawful unless it is forbidden by law or is expected to result in injury to a person or property.

Notice and consent

The PDP Bill directs data fiduciaries to notify data principals, at the time of collection of personal information, about the nature of data collected, purpose for which such data is being collected and the details of persons/entities with whom such data may be shared. It also

requires them to obtain consent from data principals before processing their personal data.

Purpose limitation

The law prescribes that data may only be processed for the specific purpose for which the data principal has consented.

Data minimisation

The law mandates that data collection shall be minimal and limited to the extent that is necessary for the purposes of processing such personal data.

Storage limitation

Data fiduciaries are barred from retaining any personal data

beyond the period necessary to satisfy the purpose for which it is processed. They are obligated to delete the personal data at the end of the processing.

Data Quality

Data fiduciaries must ensure that personal information is complete, accurate, updated and not misleading.

Accountability

The proposed framework not only prescribes measures to be undertaken to protect personal data, but it also holds data fiduciaries accountable for ensuring that the provisions are duly implemented in a demonstrable way.

Subsequent sections of the handbook seek to explain these principles with emphasis on their implementation while processing personal data.

Lawful Purpose

Are you sure that you are not processing for a purpose that is forbidden by law or could potentially cause injury/harm to an individual or community?

Notice and Consent

Have you informed the data principal about the data collected, purpose of collection and persons/entities that you share data with? Have you obtained consent for this?

Purpose Limitation

Ensure that you only process data for the purpose informed to the data principal

Data Minimisation

Ensure that you only collect the amount needed for the purpose you have stated and not more

Storage Limitation

Do you have a protocol to ensure that data is only used for the purpose

and the time stated, and data is deleted subsequently?

Data Quality

Do you have a verification mechanism to ensure that data is complete, accurate and updated?

Accountability

If you are a data fiduciary (See glossary for definition), do you ensure that the above principles are practised at data collection and processing?



Personal Data

At a glance

Understanding whether you are processing personal data is critical to understanding whether the provisions of Personal Data Protection law apply to your activities.

Personal data is information that relates to an identified or identifiable individual.

What identifies an individual could be as simple as a name or a number or could include other identifiers such as an IP address or a cookie identifier, or other factors.

If the data you are processing is capable of identifying an individual directly, it is personal data.

If an individual is not directly identifiable, then you need to consider whether the individual is identifiable by using all the means reasonably likely to be used by you or any other person to identify that individual.

It is possible that the same information is personal data for one fiduciary's purposes but is not personal data for the purposes of other fiduciary.

Introduction

The proposed data protection framework provides for greater compliance and accountability mechanisms, enforced through stricter penalties. However, the scope of the said framework is limited to personal data.

Hence, determining whether you are processing personal data or not is key to determining whether your activities fall under the ambit of data protection law.

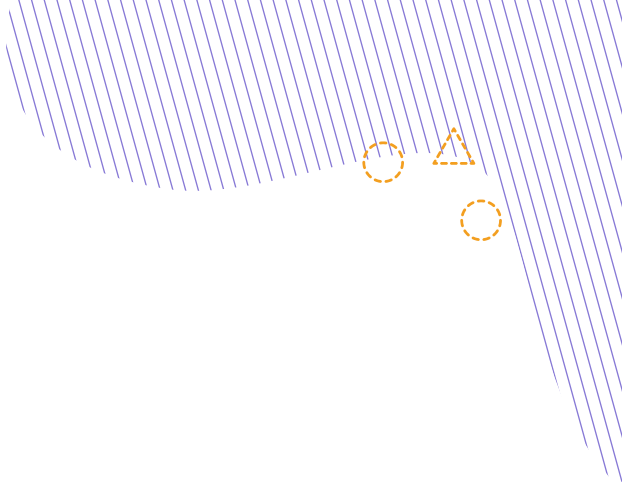
The PDP Bill defines personal data as -

data about or relating to a natural person who is directly or indirectly identifiable, having regard to any characteristic, trait, attribute, or any other feature of the identity of such natural person, whether online or offline, or any combination of such features with any other information, and shall include any inference drawn from such data for the purpose of profiling.

Based on the definition, there are two elements of personal data:

- I. It is about an individual or relates to an individual.
- II. The said data can directly or indirectly identify the individual.

In this chapter, we seek to break down the definition of personal data into individual elements



What is Personal Data?

In general terms, personal data includes any information about an individual which differentiates her/him from the rest of the world. It may include a variety of information about an individual like name, address, phone number, blood type, photos, fingerprint, bank account number, etc.

Under PDP Bill, the standard of determining whether data is personal relates to an identified or identifiable natural person.

An individual is said to be 'identified' or 'identifiable' if he/she can be distinguished from other individuals. Hence, any data that can either be directly or indirectly used to identify a certain individual is personal data.

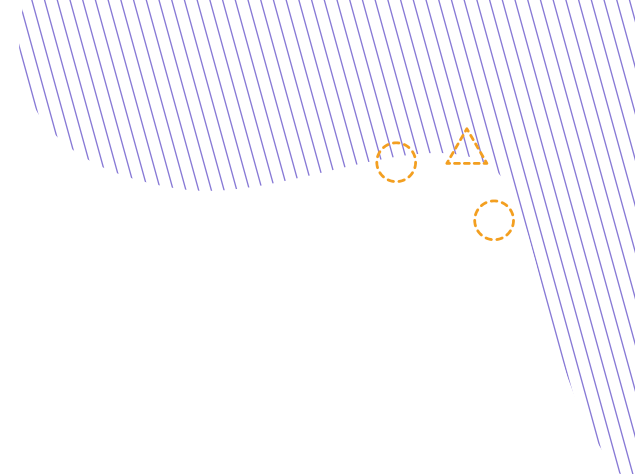
Information that can identify individuals are called identifiers. Identifiers like name, e-mail address or Aadhaar number can directly distinguish an individual from the rest. Other identifiers like

age, gender and zip code can indirectly identify an individual. Different identifiers taken together may also accurately identify an individual. The definition specifies that these identifiers can be online (information pertaining to applications, protocols, tools and devices used by an individual like IP addresses, MAC addresses, cookies, pixel tags, account handles, device fingerprints, etc.) or offline in nature.

Examples of Personal Data

Where they relate to an individual, the following data will be considered personal data:

- Name, pseudonym, date of birth
- Postal address, phone number, e-mail address, IP address
- Password, fingerprint, retinal print
- Aadhaar number, Permanent Account Number (PAN)
- Photos, video, sound recording
- Bank account details, vehicle registration number
- CCTV footage identifying a person's face



How to determine whether data 'relates to' an individual?

For the data to be considered personal, it not only has to identify an individual, but also be about or related to the individual in some way. In common terms, data relates to an individual if it is about that individual. In such cases, the relationship is easily established.

In order to determine whether data relates to an individual or not, the following factors may be considered:

a. Content of the data:

Data may relate to an individual if the underlying information is obviously about an individual. For example, medical records or criminal records of an individual. Alternatively, data may also be related to

an individual if the underlying information is not about the individual per-se but is linked to their activities. Examples of such data include itemised telephone bills or bank records.

b. Purpose of processing:

If the data is likely to be processed to learn/evaluate an individual or make a decision about an individual, then such data will be considered as personal data. Consider the phone records of a desk phone in a workplace. While the records, per se, do not relate to an individual if they are processed to ascertain the attendance of the employee manning the said desk. Such phone

records will then assume the character of personal data.

c. Effect of processing on the data principal:

If the processing of data has a resulting impact on the individual, such data will be considered personal data even if the content of the data is not directly related to the individual. For example, a factory records data regarding operation of a machinery. This data per-se does not relate to an individual. However, if the same data is processed to assess the performance of the person operating such machinery based on which his salary will be computed, such data will be considered personal data.

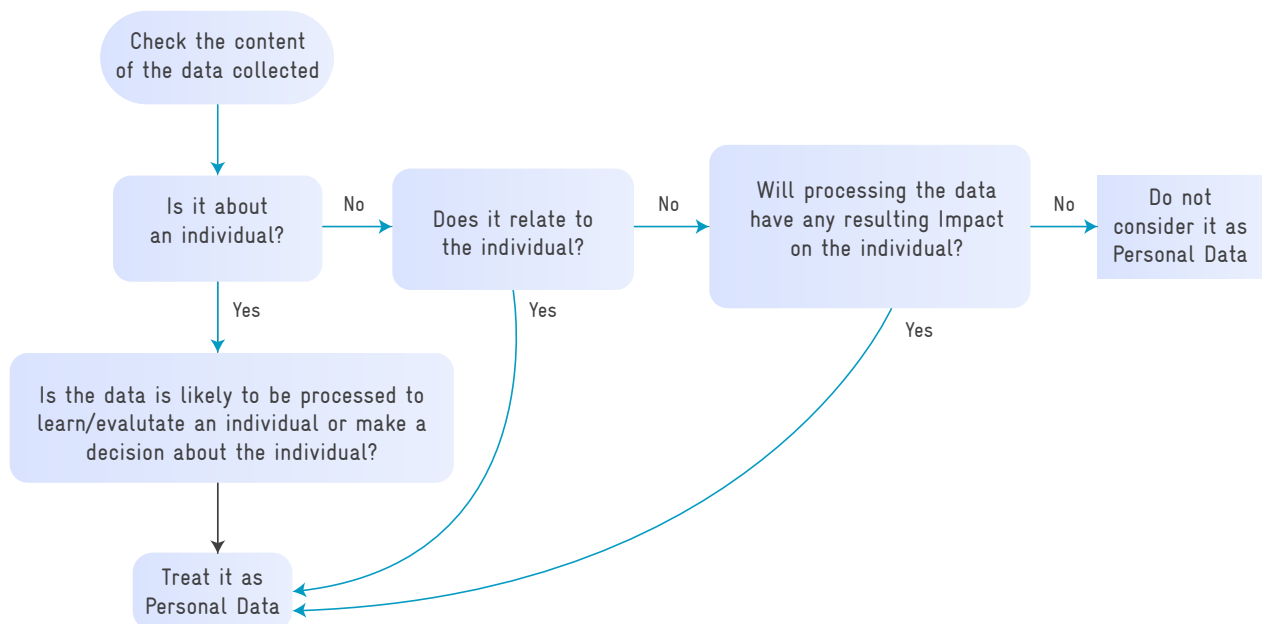
What is meant by direct or indirect identification?

If you can distinguish an individual by just looking at the information you are processing (for example, name, address, Aadhaar number), you have directly identified the individual. On the other hand, you can also identify a person using secondary information like passport number or car registration number. This sort of identification is called indirect identification. Any information, which can identify a person, whether directly or indirectly, is personal data.

How to ascertain whether the data is personal?

Please refer to the flowchart below:

Figure 6: Flowchart to assess the nature of data



What should you do if you are unable to determine whether data is personal or not?

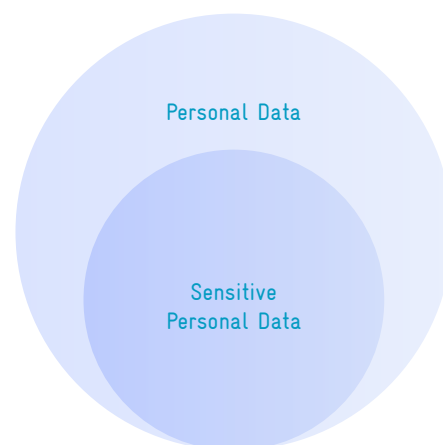
Upon assessing the data you are processing against the aforementioned metrics, if you are still unsure whether the nature of the data is personal or not, it is advisable to treat the data as personal and handle it with the highest standard of care. This will hedge you against punitive action caused by misidentification of the nature of data.

Are there other categories of personal data?

Some identifiers are classified as more sensitive than others. This class of data is known as sensitive personal data and is treated as a special category. It must be given extra protection because due to its nature, breach of such data can cause irreparable harm to the person to whom it relates. The Bill provides a non-exhaustive list of identifiers that are considered sensitive. These include - financial data, health data, biometric data, genetic data, sexual orientation, and religious or political beliefs among others.

Another categorisation provided under the Bill is that of 'critical personal data'. While the term has not been defined in the Bill, it has been accorded maximum sensitivity under law. There is prohibition on processing of critical personal data outside India.

Sensitive personal, critical personal and personal data must be held separately from each other.



Checklist for Developers

Are you able to distinguish between the different categories of data - personal data, sensitive personal data, non-personal data, etc.?

Consider creating a referencer distinguishing data into categories for the kind of data you as a startup usually deal with.

Do you understand how to distinguish data that is personally identifiable and data that is not based on content of the data, purpose of processing and effect of processing?

Do you use technical measures to secure personal data, sensitive personal data and other data that may be critical, such as anonymisation, pseudonymisation and encryption?

Data Fiduciaries and Data Processors

At a glance

Determining whether you are a data fiduciary or a data processor is critical to understand your liabilities under the PDP Bill. The Bill holds data fiduciaries liable for complying with all provisions while processors have limited liabilities.

Data fiduciaries are entities that determine the purpose and/or means of collection and processing of personal data.

Data processors are entities that process personal information on instructions of the data fiduciaries.

Introduction

The PDP Bill defines two entities that deal with personal information, viz. data fiduciary and data processor. The law prescribes varying levels of liability associated with the two

entities. Data fiduciaries shoulder the bulk of compliance requirements prescribed under law and are obligated to demonstrate their compliance of the data

protection principles. Data processors, on the other hand, have limited compliance burden. The following table maps the liabilities associated with the two entities.

Principle/Provision	Liability	
	Data Fiduciary	Data Processor
Notice and Consent	Yes	No
Security Safeguards	Yes	Yes
Data quality	Yes	No
Reporting Data Breaches	Yes	No
Maintenance of Records	Yes	No
Audit	Yes	No
Impact Assessment	Yes	No

Much like determining whether the data you are processing is personal in nature or not, it is critical to ascertain whether you are a data fiduciary or a data processor. As stated before, your compliance burden and other liabilities under law will change as per your status as a data fiduciary or data processor.



Who is a Data Fiduciary?

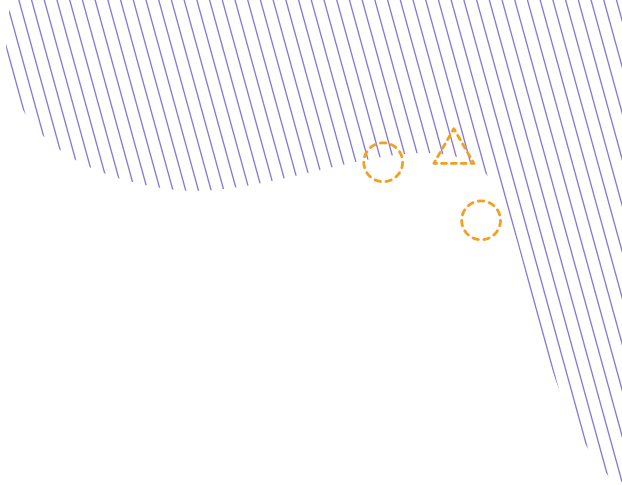
Data fiduciary is any entity (person, government agency, company/firm/trust) which determines the purpose and/or means of processing of personal data. Data fiduciaries are the main decision makers and exercise complete control over what data to process and why to process it.

For example – Bank ‘X’ appoints a technology company ‘Y’ to assess the risk of delinquency associated with certain accounts. In this case the bank is the data fiduciary as it controls the purpose and means of data processing.

Who is a Data Processor?

A data processor is an entity which processes personal data on behalf of the data fiduciary. The law defines processing as any operation performed on personal data including – collection, recording, organisation, structuring, storage, adaptation, alteration, retrieval, use, alignment or combination, indexing, disclosure by transmission, dissemination or otherwise making available, restriction, erasure, or destruction. In the example above, the company ‘Y’ hired by the bank ‘X’ becomes the data processor.

The relationship between a data fiduciary and a data processor is contractual in nature. Clause 31 mandates that data fiduciaries shall not engage a data processor without a contract in place.



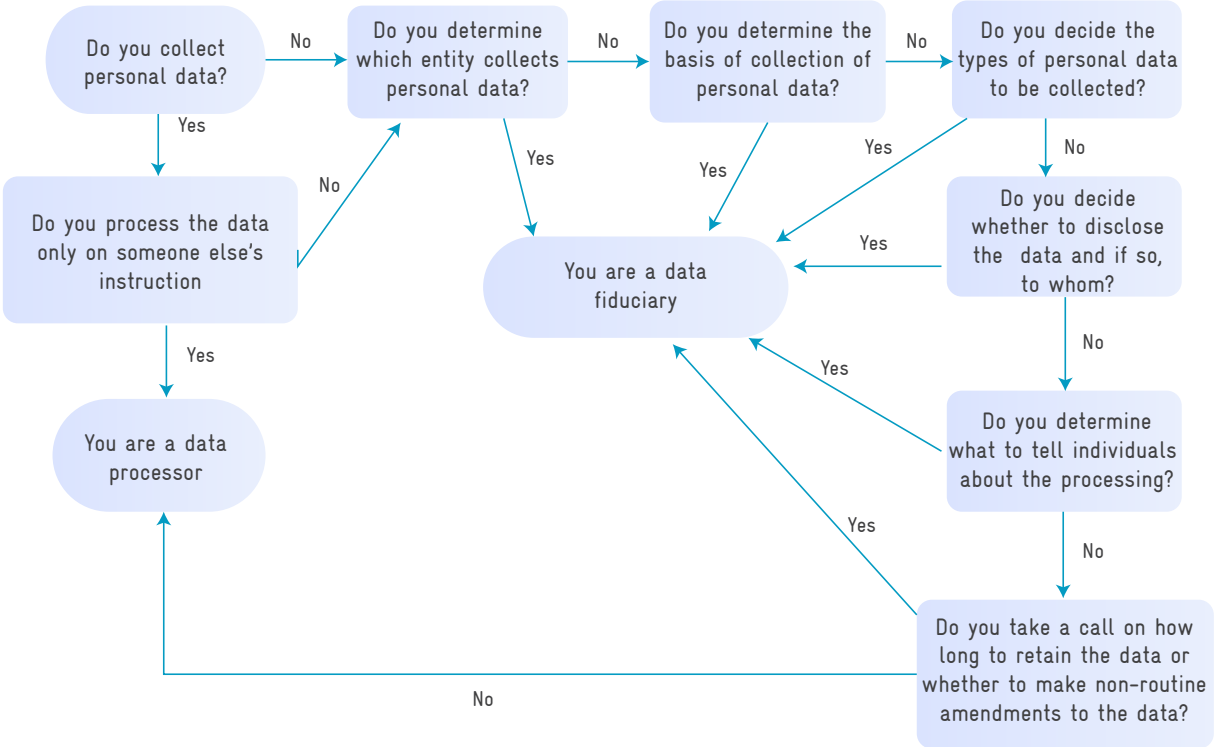
How to determine if you are a data fiduciary or a data processor?

Any entity by its very nature is not a fiduciary or a processor. To determine whether you are a data fiduciary or a data processor, with respect to a given processing activity, you need to ascertain whether you have the decision-making power regarding any of the following actions:

- Collection of personal data
- The basis for collecting personal data
- Types of personal data to be collected
- Defining the purpose for which data will be used
- Identifying the individuals about whom the data will be collected
- Whether or not to disclose the data, and if so, to whom
- What to tell individuals about the processing
- How long must data be retained to make non-routine amendments to the data

If you take a call on any or all of the aforementioned purposes and means of processing, you are a data fiduciary. If you do not determine any of the above and only process data based on someone else's instructions, you are a data processor.

Figure 7 : Steps to determine whether you are a data fiduciary or a data processor





Notice and Consent

At a glance

Informed consent is at the heart of India's data protection regime.

Data fiduciaries have to notify data subjects and provide them with relevant information at the time of collection of their personal information.

It is mandatory to obtain the principal's consent before their data is collected and processed. Data processing without obtaining consent of the data principal is prohibited by law and is punishable.

Such consent must be validly obtained and must be free from coercion, undue influence, misrepresentation and fraud.

Data principal has the right to withdraw his/her consent at a later stage.

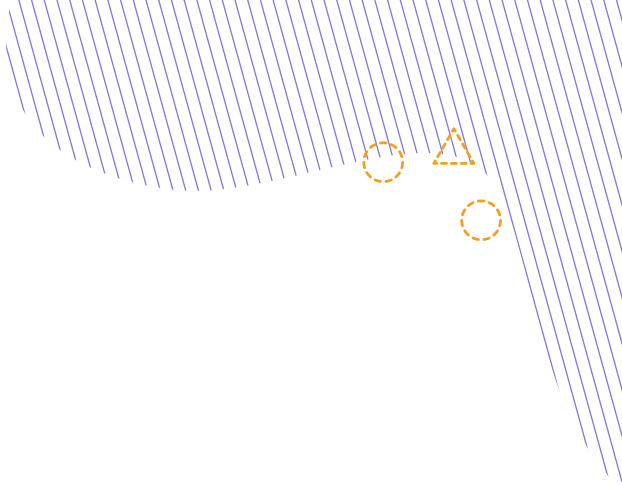
What is the principle of notice and consent?

Informed consent is at the heart of India's data protection framework. The law prohibits processing of personal data, unless the person whose data is being processed gives consent for processing. It also empowers data principals to withdraw their consent at a subsequent stage. This allows data principals to exercise control

over their personal data, while also creating scope for businesses to use an individual's personal data.

The Bill directs data fiduciaries to inform data subjects, at the time of collecting their personal information, particular details pertaining to such data collection. Among other

requirements, these include – *the purpose of data collection, the identity and contact details of the data fiduciary, the period for which such data shall be stored and persons with whom such data shall be shared. The bill also prescribes the standards of valid consent which are mentioned below:*



How do you implement the principle?

CALL to ACTION

If you are collecting personal data from an individual, you must mandatorily notify him/her about the following:

- | | |
|---|---|
| <ul style="list-style-type: none"> • the purposes for which the personal data is to be processed; • the nature and categories of personal data being collected; • the identify and contact details of the data fiduciary and the contact details of the data protection officer, if applicable, • the right of the data principal to withdraw his consent, and the procedure for such withdrawal; • the source of such collection if the | <p>personal data is not collected from the data principal;</p> <ul style="list-style-type: none"> • the individuals or entitles including other data fiduciaries or data processors, with whom such personal data may be shared, if applicable; • information regarding any cross-border transfer of the personal data that the data fiduciary intends to carry out, if applicable; • the period of which the personal data shall be retained. |
|---|---|

What is the standard of valid consent?

Clause 11(2) of the PDP Bill clarifies that consent will be considered valid only if:

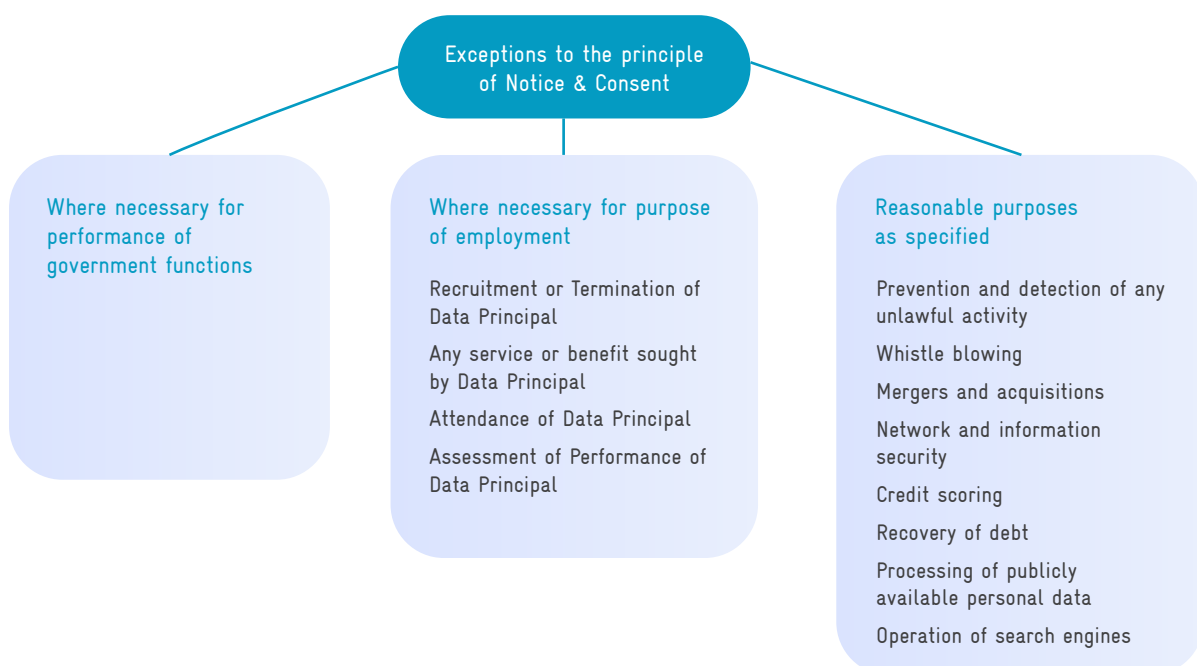
- It is obtained out of free will of the data principal i.e., the data principal was not forced or misguided to give consent.
- It is informed, i.e., the data principal has been made

aware of the nature of data collected, purpose of collection, name of entities with whom the data may be shared, period of retention, process for withdrawal of consent, etc., through a proper notice as mandated by Clause 7.

- It is specific, i.e., the data principal should be able to determine the scope of consent in respect of the purpose of processing.
- The consent is capable of being withdrawn by the data principal.

Are there exceptions to the principle of notice and consent?

Figure 8 : Exceptions to the principles of notice and consent



Checklist for Developers

**What personal data do you hold?
Where did it come from?
Who do you share it with?
What do you do with it?**

Have you identified the lawful basis for processing data and documented this?

Do you have a protocol to review how you ask for data and record consent from data principals?

While notifying data principals, have you or your organisation ensured that the notice is clear, concise, easy to understand and provides the data principal with all relevant information as described under Clause 7 of the PDP Bill? Examples of information notices can be seen [here](#).



Purpose Limitation

At a glance

The purpose for which you are processing personal data must be clear at the outset and must be communicated to the data principal.

Personal data collected can only be processed for the specified purpose.

Use of personal data for unspecified purposes is not completely prohibited. Personal data may be used for purposes incidental to or connected with the specified purpose. It may also be used for activities which the data principal may reasonably expect their data to be used for.

What is the meaning of purpose limitation?

The principle of purpose limitation is given under Clause 5(b) of the PDP Bill, which states:

Every person processing personal data of a data principal shall process such personal data

(b) for the purpose consented to by the data principal or which is incidental to or connected with such purpose, and which the data principal would reasonably expect that such personal data shall be used

for, having regard to the purpose, and in the context and circumstances in which the personal data was collected.

In simpler terms, the principle states that personal data collected for one purpose should not be processed for a different purpose. It is key to note that the law leaves a certain wiggle room for data processors in this regard by allowing processing for purposes incidental to the primary purpose for which consent was obtained.

However, processing of personal data which is completely unrelated to the stated purpose is prohibited.

According to the law, the purpose for which data is being collected must be specified at the time of collection of such data. Specifying the purpose at the outset allows data principals to furnish informed consent. It also helps data fiduciaries to remain accountable and avoid function creep.



How do you specify your purpose?

The purpose of data collection should be specified before collecting the personal information of an individual. The purpose of data collection is a part of the mandatory notification requirement specified under Clause 7 of the PDP Bill.

CALL to ACTION

You are mandatorily required to specify the purpose for which the personal data of an individual is being collected, before collecting such data.

Can the data be used for purposes other than the ones specified?

The law does not ban the use of personal data for unspecified purposes altogether. Clause 5 clarifies that personal data may be used for purposes incidental to or connected with the purpose specified. Activities which the data principal may reasonably expect their data to be used for are covered within the ambit of permitted purpose. In simpler terms, personal data may be used for purposes compatible with the specified purpose.

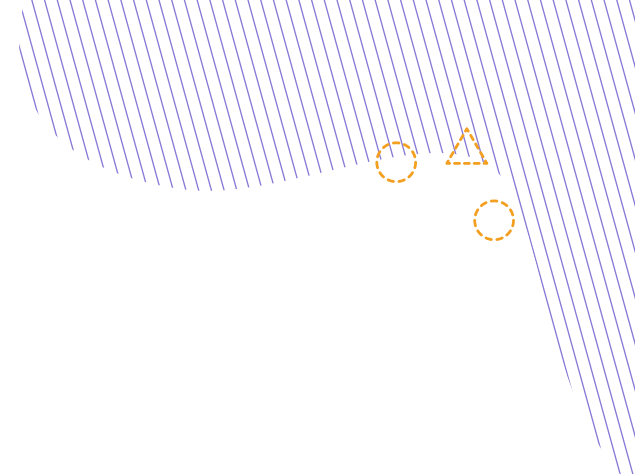
In order to assess the compatibility of purpose

vis-a-vis the originally specified purpose, you may consider the following:

- **Link between the original purpose and the new/upcoming purpose** – In simple terms, greater the distance between the originally specified purpose and the new purpose, weaker the link between the two and consequently, the more problematic it is in terms of compatibility.
- **Context in which the personal data has been**

collected – You may need to look at the nature of the relationship between the data fiduciary and the data principal. What you are expected to assess is the balance of power between the two, based on the consideration of what would be the commonly expected practice in the given relationship.

- **Nature of the personal data** – Generally, the more sensitive the information involved, the narrower the scope for compatible use.



- **Possible consequences for data principal** – You should assess the possible impact (both positive and negative) the subsequent processing may have on the data principals. Generally, risk of detriment to data principal is inversely proportional to compatibility of purpose, i.e., greater the risk of negative impact on the data principal, lesser are the chances of the purpose being compatible.
- **Existence of appropriate safeguards** – Appropriate security safeguards like encryption and pseudonymisation can offset deficiencies in other criteria

mentioned above. For example, safeguards ensure safer processing and mitigate risk of negative impact of processing on data principals.

A bank collects financial records of a person to assess his credit worthiness and offer him a personal loan. Next year, the bank processes the same data for assessing his eligibility for increased credit, without seeking his consent for the subsequent processing. This processing is allowed because the new purpose is compatible with the original purpose.

However, if the purpose is incompatible with the specified

purpose, processing of personal information for such purposes is prohi

The same bank wants to share the client's data with insurance firms. That processing isn't permitted without the explicit consent of the client as the purpose isn't compatible with the original purpose for which the data was processed.

Checklist for Developers

Do you have a clear understanding of the purpose for which data may be used?

Do you have a mechanism to ensure that this purpose is clearly represented to principals you collect data from?

If there is a change in how you use data, do you have a procedure to inform 'data principals' about this?



Data Minimisation

At a glance

Personal information should be processed only when absolutely necessary.

You must ensure that the personal data you are processing is directly relevant and necessary for achieving the stated purpose.

Relevance and necessity are not defined by law. It varies as per purpose. Hence clarity of purpose is the key to determining relevance and necessity.

What is the principle of data minimisation?

The principle is enshrined in Clause 6 of the PDP Bill. It states that -

The personal data shall be collected only to the extent that is necessary for the purposes of processing of such personal data.

Data minimisation refers to the practice of limiting the collection and processing of personal information to what is directly relevant and necessary

to achieve the specified purpose. In other words, personal information should only be processed in situations where it is not feasible to carry out the processing in another manner.

A study of businesses in the UK, France and Germany revealed that 72 per cent respondents gathered data they did not subsequently use⁷.

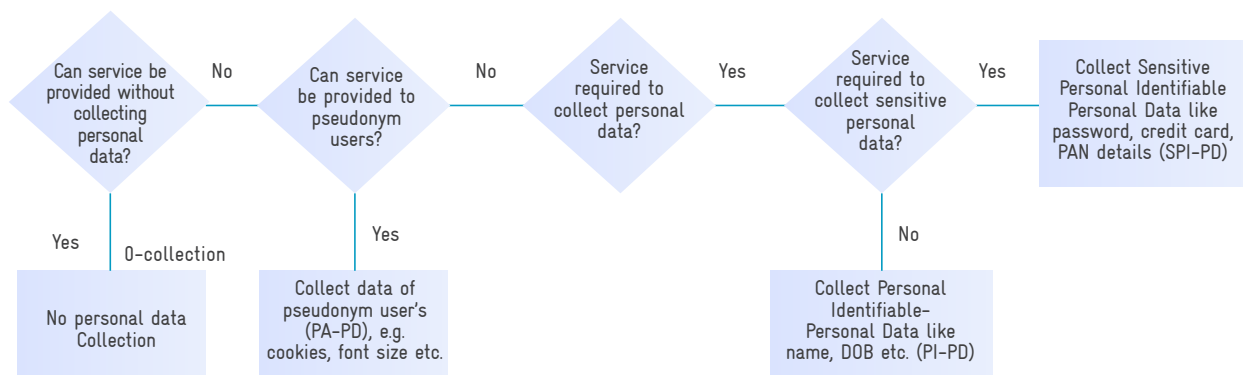
⁷ http://info.purestorage.com/rs/225-USM-292/images/Big%20Data%27s%20Big%20Failure_UK%281%29.pdf?atlid=64921319

How do you determine what is relevant and necessary?

The law does not define the parameters or threshold of relevant and necessary personal information. The relevance and necessity depend on the purpose for which information is being collected. For example, consider a database consisting of name, age, gender, educational qualification and caste of certain individuals. Now, if a company uses the database to shortlist candidates for a job position, fields like name, age and educational qualification are relevant and necessary but gender and caste information are not. On the other hand, if the database is used for identifying people for vaccination, only name and age remain relevant, while educational qualifications, caste, and gender become irrelevant.

The following flowchart shows how a data fiduciary can decide what to collect and what not to collect.

Figure 9 : Data Collection Minimisation using Flow Chart



Source: A Technical Look at the Indian PDP Bill

*PA-PD means pseudonym associated personal data

*PI-PD means personal identifiable personal data

*SPI-PD means sensitive personal identifiable personal data



CALL to ACTION

Clearly determine your purpose in order to determine whether you are holding an adequate amount of personal information

Periodically assess whether the personal information you hold is necessary and relevant to your purpose. If not, it is advisable to delete the data you no longer needed.

Is data minimisation principle antithetical to AI?

Artificial intelligence and machine learning are data intensive technologies. Especially in the case of AI, the accuracy of an algorithm is often directly proportional to the volume of data it is trained on. Hence, at the first glance, application of the principle of data minimisation may seem

particularly challenging. However, that may not be the case. Data minimisation does not bar processing of large volumes of data. It only mandates that the data must be directly relevant and necessary for the purpose for which it is processed. Further, the principle is only applicable to

‘personal data’ and not non-personal data. The spirit of the law is to protect the privacy of individuals. Hence developers can always use technological methods to protect the privacy of individuals in order to ensure that they do not violate the law.

Are there legal provisions protecting big-data applications?

The PDP Bill contains a provision through which AI or other big-data applications may be exempted from the ambit of the principles of data protection. Clause 38 of the Bill authorises relevant authority to exempt certain classes of research, archiving, or statistical purposes, where processing of personal information is necessary, from the application of any of the provisions of the Bill. In determining such exceptions,

the authority must be satisfied that:

1. The compliance with the provisions of the law shall disproportionately divert resources from such purpose;
2. The purposes of processing cannot be achieved if the personal data is anonymised;
3. The data fiduciary has carried out de-identification in accordance with the code of practice specified under

section 50 and the purpose of processing can be achieved if the personal data is in de-identified form;

4. The personal data shall not be used to take any decision specific to or action directed to the data principal; and
5. The personal data shall not be processed in the manner that gives rise to a risk of significant harm to the data principal.

What data minimisation techniques are available for AI systems?

Supervised Machine Learning algorithms rely on large amounts of data. This data is used in two phases -

- a. Training phase** - where data is used to create a machine learning algorithm; and
- b. Inference phase** - where a trained machine learning algorithm is used to make predictions.


At each stage, distinct techniques exist to either minimise the use of personal data or ensure that privacy of individuals is preserved.

Training Phase

- *Feature Selection* – This involves segregation of relevant features from irrelevant features. A dataset may contain a large number of features or data fields - like name, age, gender, pin code. However, not all features of a dataset may be relevant to your purpose. Consider a dataset containing features like name, age, gender, address, blood group, sexual orientation and political affiliation. If this data set is to be used to train a ML

model for assessing credit risk, features like sexual orientation, political affiliation and blood group are irrelevant. In order to segregate relevant features from the irrelevant features, standard methods of feature selection may be used.

- *Privacy Preserving Methods* – A range of privacy preservation techniques may be used to minimise data processing and preserve individual privacy. These include:
- *Perturbation* - This involves adding noise to data by



modifying data points in order to ensure that information cannot be traced back to specific individuals.

- *Federated Learning* - Contrary to the standard approach of centralising training data, federated learning uses a decentralised approach where ML models are trained on edge devices (like mobile phones). In the process, the data remains on local devices, thus mitigating privacy risks.
- *Synthetic Data* - Synthetic data is artificially generated data which does not relate to real people or events. It is used as a stand-in for real datasets and possesses the same statistical properties as the dataset it replaces. Thus, when used as a training set, it performs like that real-world data would.

Inference Phase

- Converting personal data into less 'human readable' format - ML models usually require the full set of predictor

variables for an individual to be included in the query in order to make a prediction about the individual. Personal data in a query can be converted into abstract formats which are less interpretable by humans. For example, in facial recognition technology, images of the faces are converted into mathematical representations of the geometric properties of the faces. These are called 'faceprints'. Instead of sending the image itself to a model for prediction/classification, the image can be converted to a faceprint, making it less identifiable.

- *Making inferences locally* - Similar to federated learning, where ML models are trained on local devices, inferences may also be made locally. This can be achieved by hosting the ML model on the device generating the query. An example of this approach is Privad, an online advertising

system which seeks to preserve user privacy by maintaining user profile (for predicting which types of ads a user will prefer) on the user's device instead of the cloud.

- *Privacy preserving queries* - In cases where local hosting of ML models is not feasible, techniques allowing minimisation of data revealed in a query sent to a model may be explored. Techniques like Trustworthy privacy-aware participatory sensing (TAPAS) allow retrieval of prediction or classification from a model without compromising the privacy of the user generating the query.

Please refer to Annexure - A for further details on technical measures that can be deployed to ensure data minimisation and privacy preservation.

Checklist for Developers

Do you have a quality check on datasets you collect, in order to remove excess data that you may not need ?

Do you systematically delete data that you do not need anymore, even though you may have lawful consent to keep data for longer ?



Storage Limitation

At a glance

You must not keep personal data for longer than you need it.

You need to think about – and be able to justify – how long you keep personal data. This will depend on your purposes for holding the data.

You should also periodically review the data you hold, and erase or anonymise it when you no longer need it.

Individuals have a right to erasure if you no longer need the data.

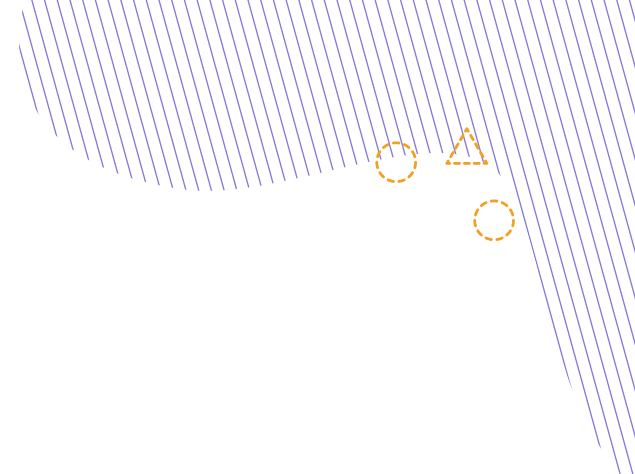
You can keep personal data for longer if it is mandated by law or is explicitly consented to by the data principal.

What is the storage limitation principle?

Even if personal information is collected and used in compliance with all the norms, personal information cannot be stored indefinitely. Personal information, which is not necessary or needed, must be deleted by data fiduciaries. This principle is enshrined in Clause 9 of the PDP Bill, which states:

- (1) The data fiduciary shall not retain any personal data beyond the period necessary to satisfy the purpose for which it is processed and shall delete the personal data at the end of the processing.
- (2) Notwithstanding anything contained in sub-section (1), the personal data may be retained for a longer period if explicitly consented to by the data principal, or necessary to comply with any obligation under any law for the time being in force.
- (3) The data fiduciary shall undertake periodic review to determine whether it is necessary to retain the personal data in its possession.
- (4) Where it is not necessary for personal data to be retained by the data fiduciary under sub-section (1) or sub-section (2), then, such personal data shall be deleted in such manner as may be specified by regulations.

Storage limitation is closely related to the data minimisation principle. Deleting personal data which is no longer necessary reduces the risk that it becomes irrelevant, excessive, inaccurate or out of date. On the practical side, storing personal data indefinitely is an inefficient practice, given the costs of storage and security of such data.



For how long can personal data be stored?

The law does not prescribe a time-limit in this regard. It merely states that personal data must not be stored beyond the period necessary to satisfy the purpose for which it is processed. Once again, the necessity of retention will depend on the purpose for which the data was collected and processed.

Consider the CCTV system installed at an ATM facility. Its purpose is to prevent fraud. Since a dubious transaction may not come to light immediately, banks need to retain the CCTV footage for a long period of time. On the other hand, a restaurant may need to retain its CCTV footage only for a shorter

period of time as any untoward incident is immediately identified.

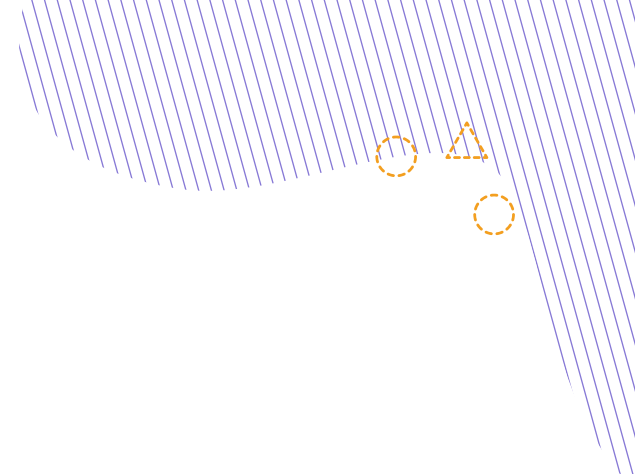
In any case, data fiduciaries are mandated to inform the data principal at the time of collecting his/her personal information as to how long his/her personal information will be stored.

When should you review the necessity of retention?

Sometimes, data retention is mandated by law. In India, laws like Companies Act, 2013, SEBI Disclosure Regulations, 2015, labour laws and Income Tax Act, 1961 prescribe data to be retained for a specified period. Such provisions will

override the necessity of purpose. In other words, if the prescribed period under any law exceeds the period for which you need to retain data, the data must be stored for the prescribed period.

Further, if the data principal consents to storage of his/her personal information for a longer period of time, such personal information may be stored beyond the period of necessity.



When can personal data be stored for a longer period of time?

The law does not prescribe the period after which data fiduciaries must review their retention. The need for retention of personal data must be reviewed at the end of any standard retention period. Unless there is a clear justification of keeping

personal data for a longer period, it must either be deleted or anonymised. In the absence of a set retention period, it is ideal to review your need of retaining personal data regularly, subject to resources and the associated privacy risks.

You are also obligated to review whether you still need personal data if an individual exercises his right to erasure. The right to erasure is discussed in detail under the chapter called Rights of Data Principals.

Checklist for Developers

Do you have a mechanism to remind you that the data retention period has lapsed or is about to lapse?

Have you considered a mechanism to seek a renewal on data retention period, in case you need data for longer than the data retention period?

Do you have a mechanism to remove data if the data principal has asked for the data to be removed/ forgotten?

Data Quality

At a glance

You should take all reasonable steps to ensure the personal data you hold is not incorrect or misleading as to any matter of fact.

You may need to keep the personal data updated, although this will depend on what you are using it for.

If you discover that personal data is incorrect or misleading, you must take reasonable steps to correct or erase it as soon as possible.

What is the principle of data quality?

In common parlance, data is considered high-quality if it is fit for its intended uses in operations, decision making and planning. From a technical perspective, quality of data depends on certain characteristics. These are:

Characteristic	How it's measured
Accuracy	Is the information correct in every detail?
Completeness	How comprehensive is the information?
Reliability	Does the information contradict other trusted resources?
Relevance	Do you really need this information?
Timeliness	How up-to-date is the information? Can it be used for real-time reporting?

The principle of data quality seeks to ensure that personal information collected and processed by data fiduciaries should be high-quality. It implies that the personal data being used should be relevant to the purpose for which it is to be used and should be accurate, complete and kept up-to-date. This is a good practice in terms of information management, but it is also linked to the rights of the individuals.

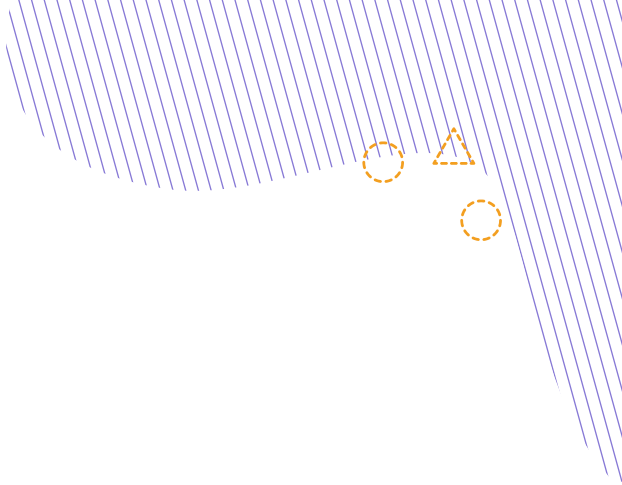
Experts suggest that “*the use of low-quality, outdated, incomplete or incorrect data at different stages of data processing may lead to poor predictions and assessments and in turn to bias, which can eventually result in infringements of the fundamental rights of individuals or purely incorrect conclusions or false outcomes*”⁸.

Further, poor data quality also poses financial challenges. Studies reveal that 60 percent of data scientists spend most of their time cleaning and sorting data.⁹ In fact, poor quality is generally regarded as the biggest obstacle to monetising data.¹⁰

⁸ https://www.europarl.europa.eu/doceo/document/A-8-2019-0019_EN.html#title4

⁹ <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>

¹⁰ <https://www.wsj.com/articles/ai-efforts-at-large-companies-may-be-hindered-by-poor-quality-data-11551741634>



When is data accurate/inaccurate?

PDP Bill does not define the meaning or parameters of accuracy of data. However, Clause 8 of the Bill specifies that accuracy, and how up-to-date the data is, can be linked to the purpose for which it is being processed. For example, let's consider an individual 'X' who moves his house from New Delhi

to Mumbai. However, in a certain dataset, X's city of residence is indicated as New Delhi. Now, if the dataset is to be used by a courier service for the purpose of delivering a product, the information is obviously inaccurate. However, if the same dataset is being used by the

Government of Delhi to identify individuals who have resided in New Delhi in the past, the data will be accurate. Hence, you must always be clear about what you intend the record of the personal data to show? What you intend to use it for may affect whether it is accurate or not.

Who is responsible for maintaining data-quality?

Personal data is intrinsically linked to individuals, who are therefore the most reliable source of data. The primary responsibility to provide accurate data to the data fiduciary ideally rests on the data principal. However, there is a corresponding obligation to ensure that data is complete, i.e., it will satisfy the purpose for which it was collected on the data fiduciary who is collecting such data. In general, it is a good practice to ensure that the information as well as its source is accurately captured. Clause 8

holds data fiduciaries responsible for maintaining the quality of personal data being processed. It states that data fiduciaries must take necessary steps to ensure that the personal data processed is complete, accurate, not misleading and updated. However, it does not define what these necessary steps must be. What constitutes a 'necessary step' will depend on the nature of the data and the purpose for which it is being used. If you are using the data for a purpose where the accuracy of the data is of

essence, you must put in more effort to ensure that the data is accurate. In cases where data is used to make decisions that may significantly affect the individual concerned or others, ensuring accuracy will require extra efforts. For example, if data is being collected for the purpose of contact tracing in order to arrest the spread of an infectious disease, ensuring the accuracy of information like names and addresses becomes critical.

Checklist for Developers

Have you considered a mechanism to filter data at the stage of receipt of such data and remove data that is incomplete, inaccurate or unnecessary?

Do you take proactive steps to account for and address the margin of error or inaccuracy in the data you collect?

Do you have a protocol for quality control if you are not the entity collecting data?

Do you have a protocol for quality control over public data and/or government data?



Accountability and Transparency

At a glance

Data fiduciaries are responsible for complying with the provisions of the PDP Bill.

Accountability goes beyond mere compliance and includes demonstrability of such compliance.

Accountability obligations are ongoing. You must review and, where necessary, update the measures you put in place.

There are a number of measures that can be taken:

- Taking a ‘data protection by design’ approach
- Implementing appropriate security measures
- Recording and, where necessary, reporting personal data breaches
- Carrying out data protection impact assessments for uses of personal data that are likely to result in high risk to individuals’ interests
- Appointing a data protection officer

Introduction

Data protection laws are designed to protect individuals from the excesses of this power imbalance. Notice and consent were the traditional instruments used by law to achieve this objective. It offered the individual the autonomy to decide whether or not to allow their data to be processed after providing their full knowledge of what was going to be done with that data. However, abundance of online services and the increasing complexity of data use cases have made

notices lengthier and more complex, making it difficult for users to comprehend. This weakens the purpose of informed consent. To offset these issues, key principles that have emerged are that of accountability and transparency.

While accountability is in-built in extant legal provisions concerning data governance, the PDP Bill takes it a step ahead. It mandates businesses to take a proactive, systematic

and answerable attitude toward data protection compliance. In this sense, accountability shifts the focus of privacy governance to an organisation’s ability to demonstrate its capacity to achieve specified privacy objectives. Thus, there are two discernible elements of accountability under the proposed framework. Firstly, it makes clear that you (data fiduciary) are responsible for complying with the law; and secondly, you must be able to demonstrate your compliance.

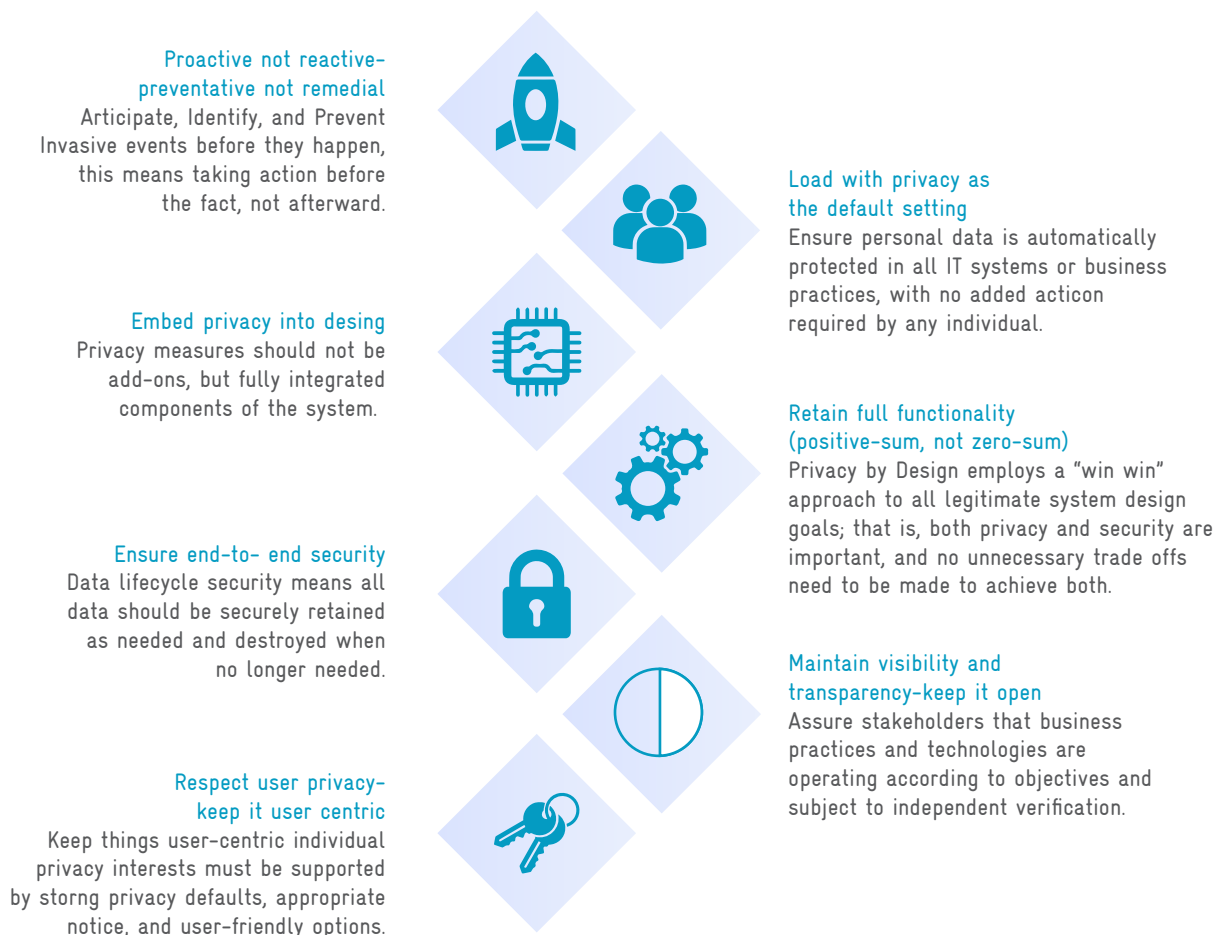
Privacy by Design

Central to this new approach to accountability is the concept of ‘Privacy by Design’.

Privacy by design as a concept was designed in the 1990s by Dr. Ann Cavoukian. It advanced the view that privacy assurance cannot be achieved by mere compliance with regulatory framework. Rather privacy must become an organisation’s default mode of operations. Initially it was limited to embedding privacy enhancing technologies (PETs) into the design of information technologies and systems. Subsequently, it evolved into an all-encompassing approach extending to a “Trilogy” of encompassing applications:

1) IT systems; 2) accountable business practices; and 3) physical design and networked infrastructure¹¹. Privacy by Design entails seven foundational principles, which are:-

Figure 10 : Seven foundational principles of Privacy by Design



¹¹ <https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf>



How would Privacy by Design be Implemented by Law?

Clause 22 of the PDP Bill mandates every data fiduciary to prepare a privacy by design policy. The policy is supposed to describe the managerial, organisational, business practices and technical systems designed to anticipate, identify and avoid harm to the data principal. It should also describe the ways in which privacy is being protected throughout the stages of processing of personal data, and how the interest of the individual is accounted for in each of these stages.

Data fiduciaries are expected to get their privacy by design policies certified by the Data Protection Authority. Once certified, the law mandates that the policy be published on the website of the data fiduciary.

Additional Resources

The following resources can be referred to by organizations and developers to further understand and implement privacy by design:

- Norwegian Data Protection Authority's [guidance](#) on how software developers can implement data protection by design.
- A [primer on privacy by design](#) published by the Information & Privacy Commissioner of Ontario.
- [Guidance on operationalizing privacy by design](#).



What are accountability measures prescribed by law?

The accountability measures mandated under PDP Bill are -

- **Identifying accountable entity:** Clause 10 of the Bill assigns the liability of complying with the provisions of the law to the data fiduciaries.
- **Appointing data protection officer:** Clause 30 of the Bill mandates significant data fiduciaries to appoint a Data Protection Officer for overseeing the organisation's data protection strategy and implementation.
- **Mandated security safeguards:** Clause 24 mandates data fiduciaries and data processors to implement technological security measures, including encryption and anonymisation, to preserve the confidentiality and integrity of personal data. The provision also mandates periodic review of such safeguards.
- **Data protection impact assessment:** Clause 27

mandates significant data fiduciaries to undertake a 'data protection impact assessment' before undertaking any processing involving new technologies or large-scale profiling or use of sensitive personal data such as genetic data or biometric data, or any other processing which carries a risk of significant harm to data principals. The DPIA must describe the proposed processing operation, the purpose of processing and the nature of personal data being processed. It must also assess the potential harm that may be caused to the data principals, and describe the measures for managing, minimising, mitigating or removing such risk of harm. The DPIA must then be submitted to the concerned authority, which may cease the processing or issue additional guidelines, if needed.

- **Maintenance of Records:** Clause 28 mandates significant data fiduciaries to

maintain accurate records of operations throughout data's lifecycle. The provision also mandates keeping records of the periodic assessment of security safeguards under clause 24 and data protection impact assessments under clause 27.

- **Audit requirements:** Clause 29 mandates significant data fiduciaries to undergo an annual audit of their policies and conduct of their data processing operations. The audit will particularly evaluate the clarity and effectiveness of notices, effectiveness of privacy by design policy, security safeguards in place and the level of transparency maintained by the data fiduciary. Such an audit must be carried out by an independent data auditor. Post audit, auditors may assign a rating in the form of a data trust score to the data fiduciary.

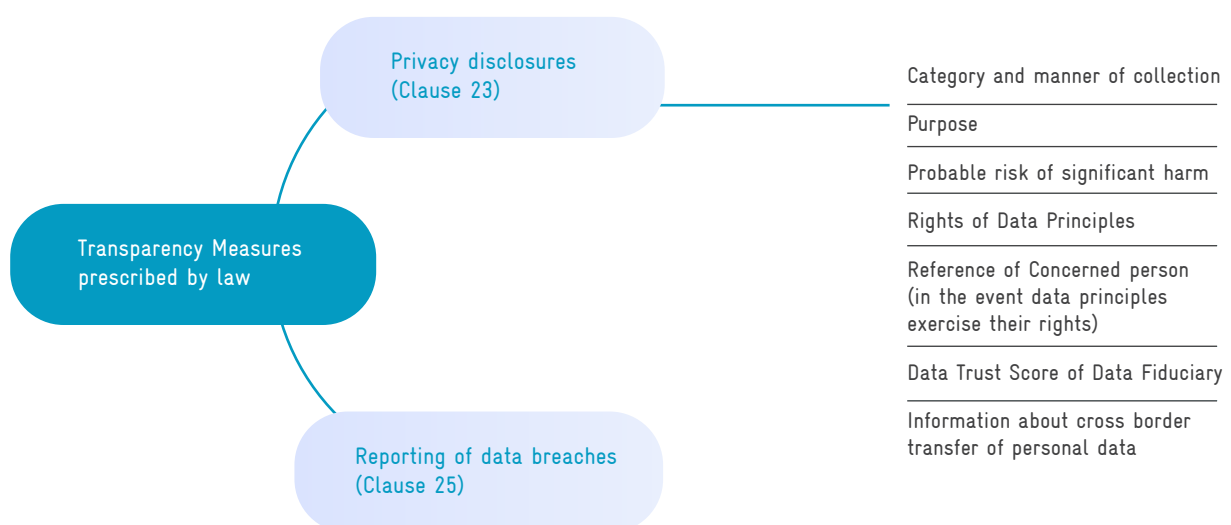
What are the transparency requirements prescribed by law?

Transparency is fundamentally linked to fairness. Transparent processing is about being clear, open and honest with people from the start about who you are, and how and why you use their personal data.

Transparency measures mandated under the PDP Bill are:

- Privacy disclosures: Clause 23 mandates every data fiduciary to disclose the following information at large:
 - a. The category of personal data collected and manner of such collection.
 - b. Purpose for which personal data is generally processed.
 - c. Data processing operations posing risk of significant harm.
 - d. Rights of data principals and the manner in which such rights can be exercised.
 - e. Contact details of persons responsible for facilitating exercise of the rights of data principals.
 - f. Data trust score earned by the data fiduciary after audit.
 - g. Information regarding cross-border transfers of personal data that the data fiduciary generally carries out (if applicable).
- Reporting of data breaches: Data fiduciaries are mandated to report, as soon as possible, any personal data breach to the relevant authority. Clause 25 states that data fiduciaries must notify the authorities about the nature of personal data breached, number of data principals affected, possible consequences, and action being taken by the data fiduciary to remedy the breach.

Figure 11 : Transparency measures under PDP Bill





CALL to ACTION

If you are a data fiduciary and there is any personal data breach, you are mandated to notify the authorities about the following:

- Nature of personal data breached
- Number of data principals affected
- Possible consequences
- Action being taken by the data fiduciary to remedy the breach

Checklist for Developers

Are you aware of your organisation's internal data protection policy?

Do you have technical mechanisms to monitor compliance with data protection policies, and regularly review the effectiveness of data handling and security controls?

Does your startup conduct Data Protection Impact Assessments? Is there a protocol for this?

Have you nominated an administrative officer to implement data protection principles or a Data Protection Officer?

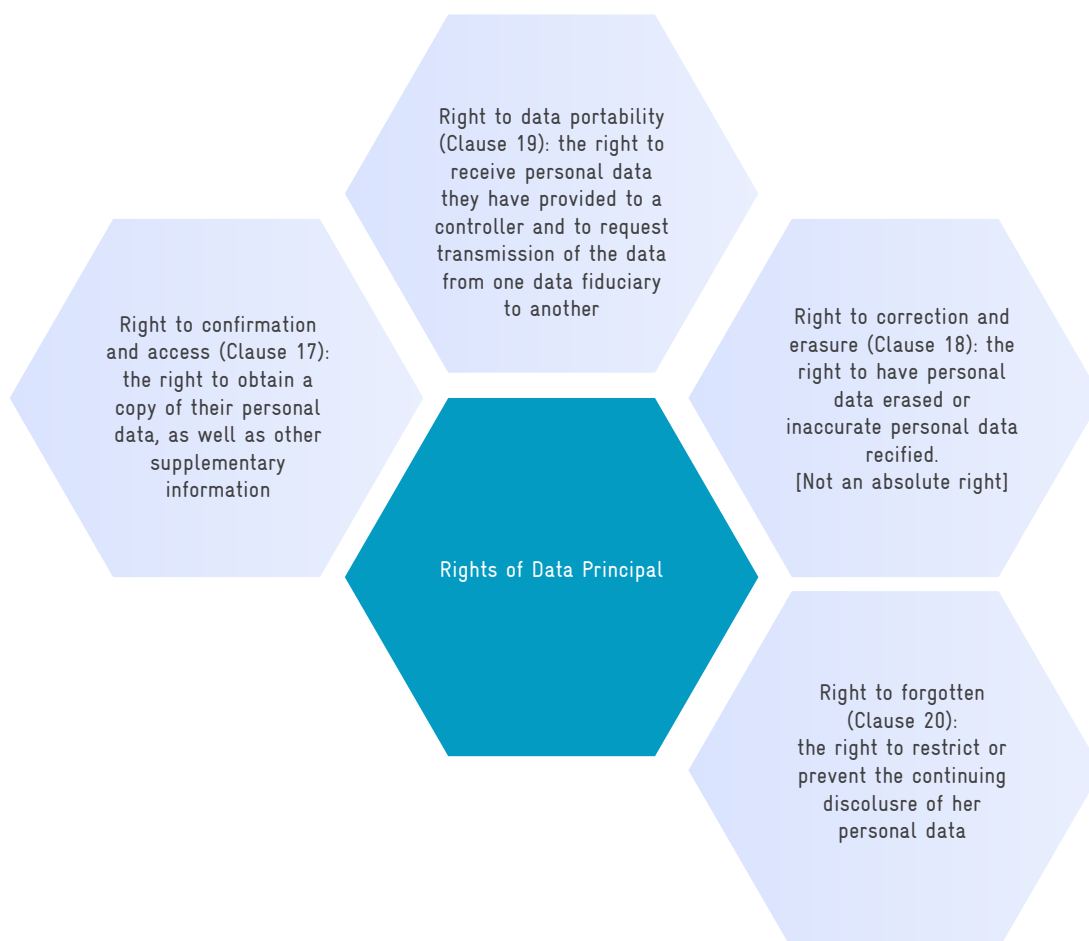
Has your startup implemented technical and organisational measures to integrate data protection practices at each stage of the processing activity?

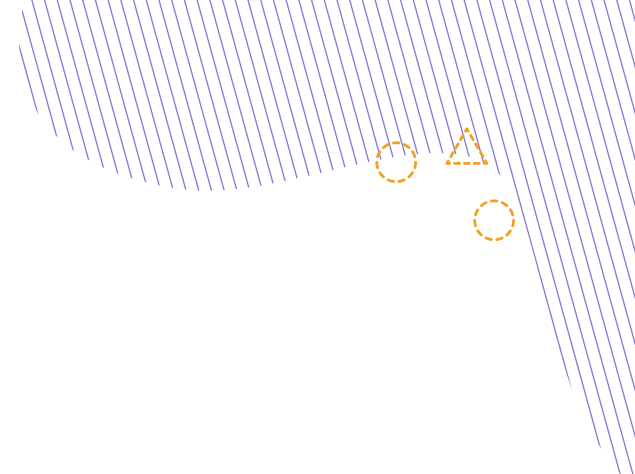
Rights of Data Principals

One of the key objectives of the PDP Bill is to empower individuals, by allowing them a greater degree of control over their personal information, to protect their fundamental right to privacy. Towards this, the Bill guarantees certain statutory rights to individuals.

It is important to understand the nature and scope of these rights as it corresponds to obligations attached to them, which have to be fulfilled by you, in your capacity, as data fiduciary/processor. Further, some of these rights also overlap with the principles of data protection discussed in previous chapters.

Figure 12 : Rights of Data Principal





Right to Confirmation and Access

Individuals have the right to obtain a copy of all the information you have about them. They are also entitled to obtain a readable copy of their data in an understandable format. This is called the right to confirmation and access. It is contained under Clause 17 of the Bill. As the name suggests, it consists of two distinct rights:

1. Right to confirmation -

A data principal has the right to obtain a confirmation from the data fiduciary on whether his/her personal data is under processing or has been processed in the past.

2. Right to access -

The data principal is entitled to access: a) his/her data which is under processing or

has been processed; b) summary of processing activities where the data has been used; and c) identity of other data fiduciaries with whom his/her data has been shared. It helps individuals to understand how and why you are using their data, and check if you are doing it lawfully.

Consider your favourite social media platform. Let's call it ABC. ABC collected personal information, like your name, date of birth, relationship status, etc., from you at the time of creating the account. It also keeps learning about your likes and dislikes by analysing your engagement with the platform. This data is personal as it relates to you and can be used to identify you. In this

example, you are the data principal and ABC is a data fiduciary. You have the right to ask ABC whether or not your data has been processed. Processing includes any operation, ranging from collection, structuring and storage to retrieval, sharing and disposal of the data.

You can also request ABC for a copy of all the data that they have on you, including inferences drawn using such data. You are also entitled to know where your data is shared and which other entity or individual has your data. ABC will be required to supply this data on your request and must be in a format that is easy to read and understand for you.

Possible implementation example for developers

Provide a functionality to display all data relating to a person. If there is a lot of data, you can split the data into several displays. If the data is too large, offer the person to download an archive containing all his or her data.

Source: CNIL. Sheet n13, Prepare for the exercise of people's rights



Right to Correction and Erasure

Clause 18 allows individuals the right to have their inaccurate personal data rectified, incomplete personal data completed, out-of-date personal information updated and have their personal data erased if it is no longer necessary for the purpose it was collected.

The right to correction has close links to the data quality principle, as it bolsters the accuracy of personal data. As essential services like banking and insurance go online, the right to correction assumes more importance. This is because inaccurate or incomplete information can lead to denial of such services.

The right to erasure, on the other hand, allows an individual to get his data deleted, if it is no longer necessary to the purpose of its

processing. For example, an individual provides his bank account details and list of assets to a bank for a loan. However, his application was denied. Since the bank no longer needs the data, the concerned individual has the right to ask the bank to delete his data.

However, it is important to note that the right to correction and erasure is not absolute and is subject to the purpose of processing. Data fiduciaries can reject a request for correction or erasure. However, they will have to furnish an adequate justification for such refusal/disagreement in writing. If the data principal is not satisfied with the explanation provided, they may require the data fiduciary to indicate that the personal information in question has been disputed by the data principal.

If the data fiduciary makes the suggested correction/ updation/erasure, it will also have to notify all relevant entities or individuals to whom such personal data may have been disclosed.

Inaccurate Information can lead to denial of services

Let's assume that the government has organised a vaccination drive for the people born in the year 1990. An individual named Ashok is eligible, as he was born in 1990. But due to a clerical error, his Aadhar Card reflects the year of birth as 1991. In the absence of this right, Ashok would have been denied vaccination because of a clerical error. However, Ashok can exercise his right to correction and request UIDAI to rectify the error.

Possible implementation example for developers

- (i) Provide a functionality to erase all data relating to a person.
- (ii) Provide for automatic notification of processors to also erase the data relating to that person.
- (iii) Provide for data erasure in backups or provide an alternate solution that does not restore erased data relating to that person.
- (iv) Allow to directly modify data in the user account.

Source: CNIL. Sheet n13, Prepare for the exercise of people's rights



Right to Portability

The right to data portability allows individuals the right to receive their personal data in a structured, commonly used and machine-readable format. It also gives them the right to request that a fiduciary transmits this data directly to another fiduciary. While it may seem similar to the right to access, the right to data portability applies to machine readable data as opposed to human readable data that you can request under right to access.

Data portability allows individuals to seamlessly switch between different service providers. It is similar to mobile number portability, which allows you to switch

between different telecom service providers without changing your number. For example, an individual uses Hotmail as his email service provider and has created a mailing list comprising his clients. However, he now wants to migrate to Gmail. Right to portability allows him to request Hotmail to either send this data (in a machine-readable format like CSV) either to him, or to Gmail directly.

As per clause 19, the personal data to be provided to the Data Principal would consist of:

I. Data already provided by the Data Principal to the Data Fiduciary;

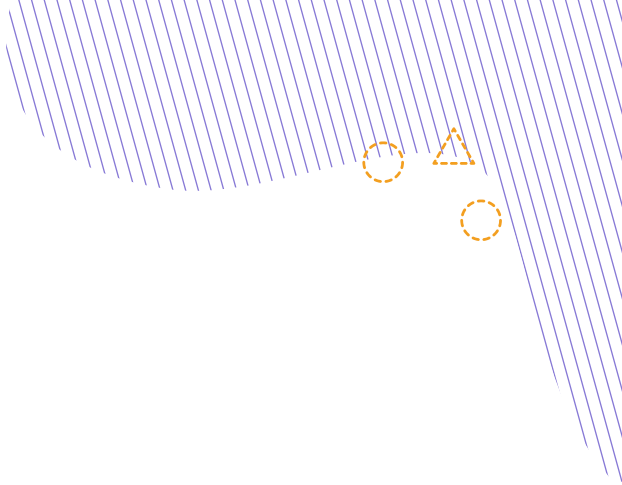
- II. Data which has been generated by the Data Fiduciary in its provision of services or use of goods;
- III. Data which forms part of any profile on the Data Principal or which the Data Fiduciary has otherwise obtained.

Exemptions have been provided for instances where (i) the data processing is not automated; (ii) where processing is necessary for compliance of law, order of a court or for a function of the State; and significantly, (iii) where compliance with the request would reveal a trade secret for a Data Fiduciary, or would not be technically feasible.

Possible implementation example for developers

Developers can provide a feature that allows the data subject to download his or her data in a standard machine-readable format.

Source: CNIL. Sheet n13, Prepare for the exercise of people's rights



Right to be Forgotten

Clause 20 of the PDP Bill pertains to the right to be forgotten. Under this Clause, the data principal is given the right to restrict or prevent the continuing disclosure of her/his personal data. It can be exercised if any one of the following conditions holds true:

- I. The disclosure of personal data is no longer necessary for the purpose for which it has been collected, or
- II. The data principal has revoked their consent to such disclosure of personal data, or

III. The disclosure of personal data is contrary to the provisions of the Bill, or any other law in force.

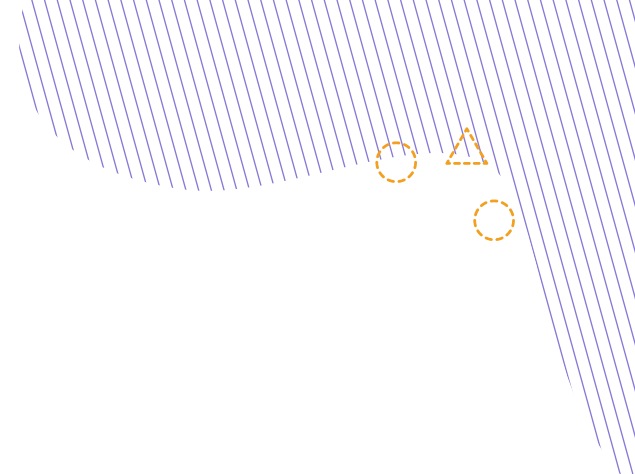
Where disclosure has taken place on the basis of the consent of a data principal, the unilateral withdrawal of such consent could trigger the right to be forgotten. In other cases, where there is a conflict of assessment as to whether the purpose of the disclosure has been served or whether it is no longer necessary, a balancing test must be carried out. The exercise of this right requires balancing the right to privacy

with the right to freedom of speech and expression.

Possible implementation example for developers

- (i) Provide a functionality allowing the data principal to restrict processing.
- (ii) If the request is approved, you must delete data already collected, and must not subsequently collect any more data related to that person.

Source: CNIL. Sheet n13, Prepare for the exercise of people's rights



How can these rights be exercised?

Except the right to be forgotten, every right under the Act can be exercised by sending a written request to the data fiduciary. Such a request can be made either directly or through a consent manager. Consent manager refers to any platform that allows individuals to manage (gain, withdraw, modify) their consent.

Upon receiving a request, the

data fiduciary has to either comply with the request or explain as to why the request was not complied with.

The right to be forgotten, on the other hand, can only be enforced upon an order of an adjudicating officer, appointed by the government. Data principals will be allowed to request the said officer to issue such an order. The officer will

determine whether or not to grant the request. However, such determination must be made considering factors like the sensitivity of the data, the scale of disclosure and the degree of accessibility sought to be restricted, the role of data principal in public life and the importance of the personal data to the public.

Checklist for Developers

Have you provided privacy information to individuals?

Do you have a process to recognise and respond to individuals' requests to access their personal data?

Do you have processes to ensure that the personal data you hold remains accurate and up to date?

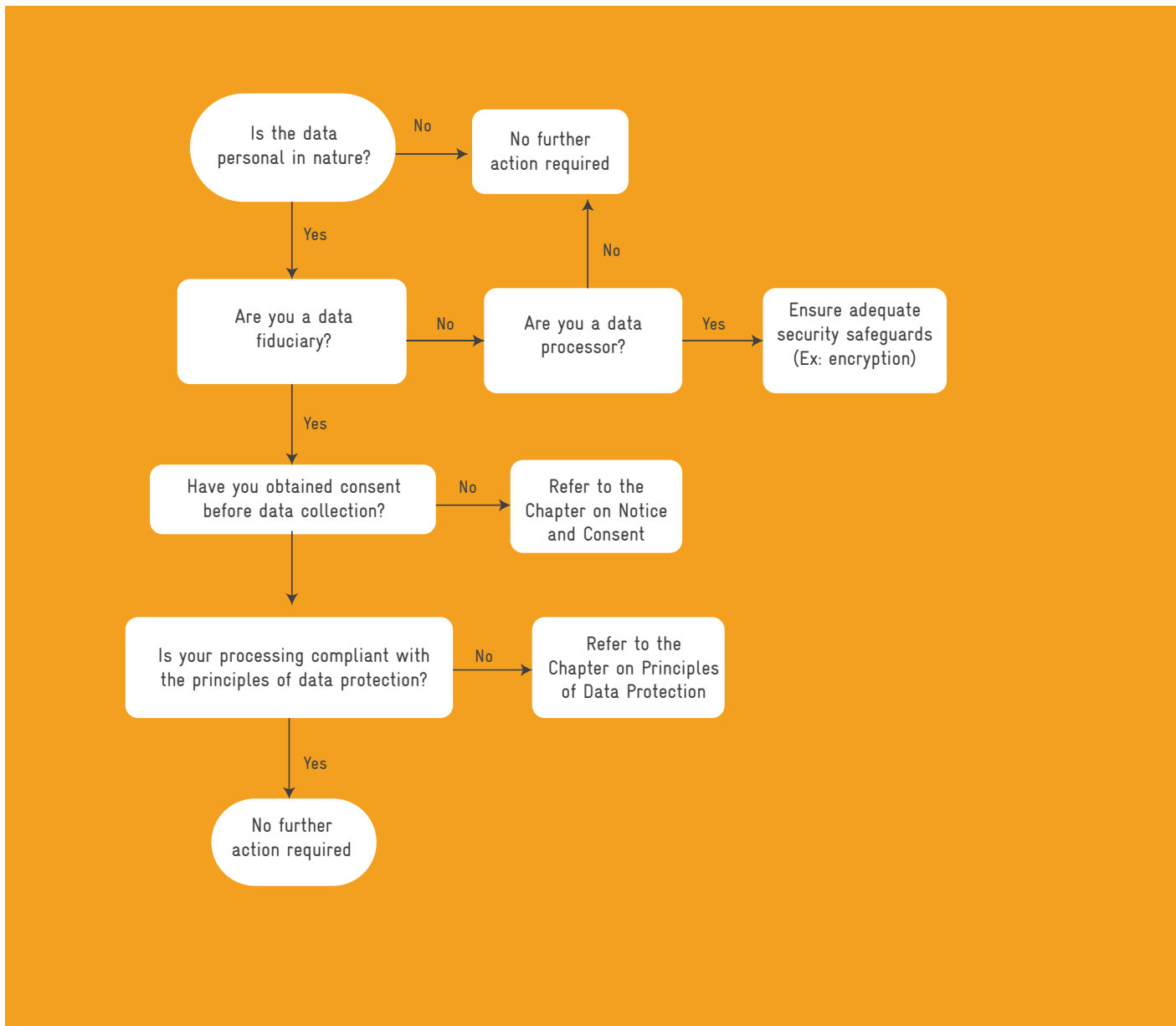
Do you have a process to securely dispose of personal data that is no longer required or where an individual has asked you to erase it?

Do you have a process to respond to an individual's request to restrict the processing of their personal data?

Do you have processes to allow individuals to move, copy or transfer their personal data from one IT environment to another in a safe and secure way, without hindrance to usability?

Compliance Map


Please refer to the following flow-chart to assess your compliance with the data protection law. The accompanying master checklist will allow you to identify actionable steps and assess the degree of your compliance.



Annexure A

Technical Measures for Data Security

Purpose	Measure	Meaning
Methods for reducing the need for training data	<u>Generative Adversarial Networks</u>	Generative Adversarial Networks (GAN) are used for generating synthetic data. As of today, GAN has mainly been used for the generation of images. But it also has the potential for becoming a method for generating huge volumes of high quality, synthetic training data in other areas. This will satisfy the need for both labelled data and large volumes of data, without the need to utilise great amounts of data containing real personal information.
	<u>Federated Learning</u>	This is a form of distributed learning. Federated learning works by downloading the latest version of a centralized model to a client unit, for example a mobile phone. The model is then improved locally on the client, on the basis of local data. The changes to the model are then sent back to the server where they are consolidated with the change information from models on other clients. An average of the changed information is then used to improve the centralized model. The new, improved centralized model may now be downloaded by all the clients. This provides an opportunity to improve an existing model, on the basis of a large number of users, without having to share the users' data.
	<u>Matrix capsules</u>	Matrix capsules are a new variant of neural networks, and require less data for learning than what is currently the norm for deep learning. This is very advantageous because a lot less data is required for machine learning.
Methods that uphold data protection without reducing the basic dataset	<u>Differential privacy</u>	Start with a database that contains natural persons and features related to these persons. When information is retrieved from the database, the response will contain deliberately generated "noise", enabling information to be retrieved about persons in the database, but not precise details about specific individuals. A database must not be able to give a markedly different result to a query if an individual person is removed from the database or not. The overriding trends or characteristics of the dataset will not change.



Purpose	Measure	Meaning
	<u>Homomorphic encryption</u>	This is an encryption method that enables the processing of data whilst it is still encrypted. This means that confidentiality can be maintained without limiting the usage possibilities of the dataset. At present, homomorphic encryption has limitations, which mean that systems employing it will operate at a much lower rate of efficiency. The technology is promising, however. Microsoft for example has published a white paper on a system that uses homomorphic encryption in connection with image recognition.
	<u>Transfer Learning</u>	It is not the case that it is always necessary to develop models from scratch. Another possibility is to utilise existing models that solve similar tasks. By basing processing on these existing models, it will often be possible to achieve the same result with less data and in a shorter time. There are libraries containing pre-trained models that can be used.
	<u>RAIRD</u>	<p>Statistics Norway (SSB) and the Norwegian Centre for Research Data (NSD) have developed a system called RAIRD. It permits research to be carried out on data without having direct access to the complete dataset.</p> <p>In short, this system works by means of an interface that allows the researchers to access only the metadata of the underlying dataset. The researcher can then submit queries based on the metadata and obtain a report containing aggregated data only. This solution has been designed to prevent the retrieval of data relating to very small groups and individual persons. This type of system can therefore be used when data for machine learning is needed. Instead of receiving a report as an end result, one could obtain a model from the system</p>

Annexure B

List of Abbreviations

ACM FAccT

Association for Computing Machinery Fairness, Accountability and Transparency

AI

Artificial Intelligence

API

Application Programming Interface

COMPAS

Correctional Offender Management Profiling for Alternative Sanctions

DPIA

Data Protection Impact Assessment

FAT Framework

Fairness Accountability and Transparency Framework

GAN

Generative Adversarial Networks

GPT

Generative Pre-trained Transformer

IEC

International Electrotechnical Commission

IP Address

Internet Protocol Address

IPR

Intellectual Property Rights

IS

International Standards

ISO

International Organisation for Standardisation

IT

Information Technology

IT Act

Information Technology Act, 2000

LIME

Local Interpretable Model Agnostic Explanations

MAC Address

Media Access Control Address

ML

Machine Learning

MNE Guidelines

OECD Guidelines for Multinational Enterprises

NSD

Norwegian Centre for Research Data

NTA

Norwegian Tax Administration

OECD

Organisation for Economic Co-operation and Development

PAN

Permanent Account Number

PA-PD

Pseudonym Associated Personal Data

PDP Bill

Personal Data Protection Bill, 2019

PETs

Privacy Enhancing Technologies

PI-PD

Personal Identifiable Personal Data

RAIRD

Remote Access Infrastructure for Register Data

SEBI

Securities and Exchange Board of India

SPDI Rules

Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011

SPI-PD

Sensitive Personal Identifiable Personal Data

SSB

Statistics Norway

TAPAS

Trustworthy Privacy-Aware Participatory Sensing

UK

United Kingdom

US

United States of America

USD

United States Dollar

XAI


Explainable AI

Annexure C

Master Checklists (Section I: Ethics in AI)

Stage of Intervention	Checklist
I. TRANSPARENCY	
Pre-Processing	<p>Are you aware of the source of data used for training?</p> <p>Have you recorded the attributes to be used for training and the weightage for each?</p> <p>Is there scope for diagnosing errors at a later stage of processing?</p>
In Processing	<p>Have you checked for any blind spots in the AI enabled system and decision-making process?</p> <p>Can the AI decision-making process be explained in simple terms that a layperson can understand?</p>
Post Processing	<p>Is it possible to obtain an explanation/reasoning of AI decisions? For instance, can a banker obtain a record for acceptance/denial of a loan application.</p> <p>Is there a mechanism for users and beneficiaries to raise a ticket for AI decisions?</p> <p>Is there scope for oversight and human review of AI decisions?</p>
II. Accountability	
Pre-Processing	<p>Does the start-up have policies or protocols or contracts on liability limitation and indemnity? Are these accessible and clear to you?</p> <p>Does your organisation have a data protection policy? Are data protection guidelines being followed in the collection and storage of data, if the data is procured from a third-party?</p> <p>Are there adequate mechanisms in place to flag concerns with data protection practices?</p>
In Processing	<p>Can the AI enabled system be compartmentalised based on who has developed the concerned portion?</p> <p>Is it possible to apportion responsibility within your set-up if multiple developers have worked on a project?</p> <p>Is it possible to maintain records about design processes and decision-making points?</p>
Post Processing	<p>Can decisions be scrutinised and traced back to attributes used and developers that worked on the project?</p> <p>Have you identified tasks that are too sensitive to be delegated to AI systems?</p> <p>Are there protocols/procedures in place for monitoring, maintenance and review of AI enabled systems?</p>

Stage of Intervention	Checklist
III. Mitigating Bias	
Pre-Processing	<p>Are you able to identify the source/sources of bias at the stage of data collection?</p> <p>Did you check for diversity in data collection before it was used as training to mitigate bias?</p> <p>Did you analyse the data for historical biases?</p>
In Processing	<p>Have you assessed the possibility of AI correlating protected attributes and bias arising as a result?</p> <p>Do you have an overall strategy (technical operational) to trace and address bias?</p> <p>Do you have technical tools to identify potential sources of bias and introduce de-biasing techniques? Please see Appendix for a list of technical tools that developers may consider.</p> <p>Have you identified instances where human intervention would be preferable over automated decision making?</p>
Post Processing	<p>Have you identified cases of where human intervention will be preferred over automated decision making?</p> <p>Do you have internal and/or third-party audits to improve data collection processes?</p>
IV. Fairness	
Pre-Processing	<p>Have you anticipated the possibility of the algorithm treating similar candidates differently on the basis of attributes such as race, caste, gender or disability status?</p> <p>Are there sufficient checks to ensure that the machine does not base its outputs on protected attributes?</p>
In Processing	<p>Are there protocols or sensitisation initiatives to reconcile historical biases and build in weights to equalise potential biases?</p> <p>Have you conducted due diligence to trace potential fairness harms that may arise directly or indirectly?</p> <p>Do you have technical tools to diagnose fairness harms and address them?</p>
Post Processing	<p>Does the algorithm provide results that are equally and similarly accurate across demographic groups?</p> <p>Have you checked outcomes to understand if the AI is producing equal true and false outcomes? Is there any scope for disproportionate outcomes?</p>



Stage of Intervention	Checklist
V. Security	
Pre-Processing	<p>Have you identified scenarios where safety and reliability could be compromised, both for the users and beyond?</p> <p>Have you classified anticipated threats according to the level of risk and prepared contingency plans to mitigate this risk?</p> <p>Have you defined what a safe and reliable network means (through standards and metrics), and in this definition consistent throughout the full lifecycle of the AI enabled system?</p>
In Processing	<p>Have you created human oversight and control measures to preserve the safety and reliability risks of the AI system, considering the degree of self-learning and autonomous features of the AI system?</p> <p>Do you have procedures in place to ensure the explainability of the decision-making process during operation?</p> <p>Have you developed a process to continuously measure and assess safety and reliability risks in accordance with the risk metrics and risk levels defined in advance for each specific use case?</p>
Post Processing	<p>Have you assessed the AI enabled system to determine whether it is also safe for, and can be reliably used by, those with special needs or disabilities or those at risk of exclusion?</p> <p>Have you facilitated testability and auditability?</p> <p>Have you accommodated testing procedures to also cover scenarios that are unlikely to occur but are nonetheless possible?</p> <p>Is there a mechanism in place for designers, developers, users, stakeholders and third parties to flag/report vulnerabilities and other issues related to the safety and reliability of the AI enabled System? Is this system subject to third-party review?</p> <p>Do you have a protocol to document results of risk assessment, risk management and risk control procedures?</p>

Stage of Intervention	Checklist
VI. Privacy	
Pre-Processing	<p>Have you considered deploying privacy specific technological measures in the interest of personal data protection? Please see the Appendix for an illustrative list of technological measures you could consider.</p> <p>Are you able to trace the source for the data being used in the AI system? Is there adequate documentation of informed consent to collect and use personal data for the purpose of developing AI?</p> <p>Is sensitive data collected? If so, have you adopted higher standards for protection of this kind of data?</p> <p>Does the training data include data of children or other vulnerable groups? Do you maintain a higher standard of protection in these cases?</p> <p>Is the amount of personal data in the training data limited to what is relevant and necessary for the purpose?</p>
In Processing	<p>Have you considered all options for the use of personal information (e.g. anonymisation or synthetic data) and chosen the least invasive method?</p> <p>Have you considered mechanisms/techniques to prevent re-identification from anonymised data?</p> <p>Have you been informed of the digital requirements for processing personal information? What measures does your organisation adopt to ensure compliance?</p>
Post Processing	<p>Are there procedures for reviewing data retention and deleting data used by the AI System after it has served its purpose? Are there oversight/review mechanisms in place? Is there scope for external/third-party review?</p> <p>Beyond the data principal' privacy, have you evaluated the potential of data of an identified group being at risk?</p> <p>Is there a mechanism in place to manage consent provided by data principals? Is this mechanism accessible to principals? Is there a method to periodically review consent?</p>

Master Checklists (Section II: Data Protection)

Checklist

Are you able to distinguish between the different categories of data - personal data, sensitive personal data, non-personal data etc.?

Consider creating a referencer distinguishing data into categories for the kind of data you as a startup usually deal with.

Do you understand how to distinguish data that is personally identifiable and data that is not - based on content of the data, purpose of processing and effect of processing?

Do you use technical measures to secure personal data, sensitive personal data and other data that may be critical - such as anonymisation, pseudonymisation and encryption?

As a start-up, have you documented:

1. What personal data do you hold?
2. Where did it come from?
3. Whom do you share it with?
4. What do you do with it?

Have you identified the lawful basis for processing data and documented this?

Does your startup have a protocol to review how you ask for data and record consent from data principals?

While notifying data principals, have you or your organisation ensured that the notice is clear, concise, easy to understand and provides the data principal with all relevant information as described under Clause 7 of the PDP Bill?

Do you provide developers with a clear understanding of the purpose for which data may be used?

Do you have a mechanism to ensure that this purpose is clearly represented to principals you collect data from?


If there is a change in how you use data, do you have a procedure to inform 'data principals' about this?

Do you have a quality check on datasets you collect, in order to remove excess data that you may not need?

Do you systematically remove data that you do not need anymore, even though you may have lawful consent to keep data for longer?

Do you train new recruits or conduct training exercises to inform employees of best practices or technical measures in data protection?

Do you have a mechanism to remind you that the data retention period has lapsed or is about to lapse?



Checklist

Have you considered a mechanism to seek a renewal on data retention period in case you need data for longer than the data retention period?

Do you have a mechanism to remove data if the data principal has asked for the data to be removed/forgotten?

Have you considered a mechanism to filter data at the stage of receipt of such data and remove data that is incomplete, inaccurate or unnecessary?

Do you take proactive steps to account for and address the margin of error or inaccuracy in the data you collect?

Do you have a protocol for quality control if you are not the entity collecting data?

Do you have a protocol for quality control over public data and/or government data?

Does your startup have an internal data protection policy?

Does your startup monitor its compliance with data protection policies and regularly review the effectiveness of data handling and security controls?

Does your startup have a written contract with any processors you use?

Has your startup implemented technical and organisational measures to integrate data protection practices at each stage of the processing activity?

Does your startup conduct Data Protection Impact Assessments? Is there a protocol for this?

Have you nominated an administrative officer to implement data protection principles or a Data Protection Officer?

Have you provided privacy information to individuals?

Does your startup have a process to recognise and respond to individuals' requests to access their personal data?

Does your startup have processes to ensure that the personal data you hold remains accurate and up to date.

Does your startup have a protocol to securely dispose of personal data that is no longer required or where an individual has asked you to erase it?

Does your startup have procedures to respond to an individual's request to restrict the processing of their personal data?

Does your startup have processes to allow individuals to move, copy or transfer their personal data from one IT environment to another in a safe and secure way, without hindrance to usability?

References

1. Solon Barocas and Andrew D. Selbst, 'Big Data's Disparate Impact' (104 California Law Review 671, 2016) https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2808263_code1328346.pdf?abstractid=2477899.&mirid=1
2. Josh Cowsls and Luciano Floridi, 'Prolegomena To A White Paper On An Ethical Framework For A Good AI Society' (2018) <http://dx.doi.org/10.2139/ssrn.3198732>
3. (Article19.org, 2021) https://www.article19.org/wp-content/uploads/2019/04/Governance-with-teeth_A19_April_2019.pdf
4. David Leslie, 'Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector' (Zenodo, 2021) <https://doi.org/10.5281/zenodo.3240529>
5. Luciano Floridi and others, 'Ethical Framework For A Good AI Society' (Eismd.eu, 2019) <https://www.eismd.eu/wp-content/uploads/2019/02/Ethical-Framework-for-a-Good-AI-Society.pdf>
6. Pranay K. Lohia and others, 'Bias Mitigation Post-Processing For Individual And Group Fairness' (arxiv.org, 2021) <https://arxiv.org/abs/1812.06135>
7. Brian d'Alessandro, Cathy O'Neil and Tom LaGatta, 'Conscientious Classification: A Data Scientist's Guide To Discrimination-Aware Classification' (Arxiv.org) <https://arxiv.org/pdf/1907.09013.pdf>
8. Ivan Bruha and A. Fazel Famili, 'Postprocessing in Machine Learning and Data Mining' (Kdd.org, 2000) https://www.kdd.org/exploration_files/KDD2000PostWkshp.pdf
9. Dr. Matt Turek, 'Explainable Artificial Intelligence (XAI)' (Darpa.mil, 2019) <https://www.darpa.mil/program/explainable-artificial-intelligence>
10. Carlos Guestrin and Marco Tulio Ribeiro, 'Local Interpretable Model-Agnostic Explanations (LIME): An Introduction' (O'Reilly Media, 2021) <https://www.oreilly.com/learning/introduction-to-local-interpretable-model-agnostic-explanations-lime>
11. Felzmann, H., Fosch-Villaronga, E., Lutz, C. et al. Towards Transparency by Design for Artificial Intelligence. Sci Eng Ethics 26, 3333–3361 (2020). <https://doi.org/10.1007/s11948-020-00276-4>
12. Alejandro Barredo Arrieta and others, 'Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI' (Arxiv.org, 2019) <https://arxiv.org/pdf/1910.10045.pdf>
13. Heike Felzmann and others, 'Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns' (SAGE Journals, 2019) <https://journals.sagepub.com/doi/full/10.1177/2053951719860542>
14. 'Putting Data At The Service Of Agriculture: A Case Study of CIAT | Digital Tools for Agriculture | U.S. Agency For International Development' (Usaid.gov, 2018) <https://www.usaid.gov/digitalag/ciat-case-study>
15. 'API Error Handling - Unique Identification Authority of India | Government of India' (Unique Identification Authority of India | Government of India, 2021) <https://www.uidai.gov.in/916-developer-section/data-and-downloads-section/11351-api-error-handling.html>
16. (Niti.gov.in, 2018) https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf
17. 'Transparency and Explainability (OECD AI Principle) - OECD.AI' (Oecd.ai, 2021) <https://oecd.ai/dashboards/ai-principles/P7>
18. Finale Doshi-Velez, Mason Kortz and Ryan Budish, 'Accountability of AI under the Law: The Role Of Explanation' (Arxiv.org, 2021) <https://arxiv.org/pdf/1711.01134.pdf>
19. Genie Barton and Nicol Turner-Lee, 'Algorithmic Bias Detection And Mitigation: Best Practices And Policies To Reduce Consumer Harms' (Brookings, 2021) <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/#footnote-7>

- 
- 
20. Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women' (U.S., 2018) <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>
 21. Point No. 12 'REPORT On a Comprehensive European Industrial Policy on Artificial Intelligence and Robotics' (Europarl.europa.eu, 2018) https://www.europarl.europa.eu/doceo/document/A-8-2019-0019_EN.html#title4
 22. 'Reducing Bias from Models Built on the Adult Dataset Using Adversarial Debiasing' (Medium, 2020) <https://towardsdatascience.com/reducing-bias-from-models-built-on-the-adult-dataset-using-adversarial-debiasing-330f2ef3a3b4>
 23. 'The Allegheny Family Screening Tool' (Alleghenycounty.us, 2017) <https://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx>
 24. Adam Hadhazy, 'Biased Bots: Artificial-Intelligence Systems Echo Human Prejudices' (Princeton University, 2017) <https://www.princeton.edu/news/2017/04/18/biased-bots-artificial-intelligence-systems-echo-human-prejudices>
 25. (2021) https://www.ftc.gov/systems/files/documents/public_events/313371/bigdata-slides-sweenezyang-9_15_14.pdf
 26. Arvind Narayanan, 'Tutorial: 21 Fairness Definitions And Their Politics' (2018) <https://docs.google.com/document/d/1bnQKzFAzCTcBcNvW5tsPuSDje8WWWY-SSF4wQm6TLvQ/edit>
 27. Sarah Bird and others, 'Fairlearn: A Toolkit for Assessing and Improving Fairness in AI' (Microsoft.com, 2020) https://www.microsoft.com/en-us/research/uploads/prod/2020/05/Fairlearn_WhitePaper-2020-09-22.pdf
 28. David De Cremer, 'What Does Building A Fair AI Really Entail?' (Harvard Business Review, 2020) <https://hbr.org/2020/09/what-does-building-a-fair-ai-really-entail>
 29. Natasha Lomas, 'Accenture Wants to Beat Unfair AI with a Professional Toolkit' (Social.techcrunch.com, 2018) <http://social.techcrunch.com/2018/06/09/accenture-wants-to-beat-unfair-ai-with-a-professional-toolkit/>
 30. (Query.prod.cms.rt.microsoft.com) <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4t6dA>
 31. HemaniSheth, '52% Of Indian Organisations Suffered A Successful Cyber security Attack In The Last 12 Months: Survey' (@businessline, 2021) <https://www.thehindubusinessline.com/news/52-of-indian-organisations-suffered-a-successful-cybersecurity-attack-in-the-last-12-months-survey/article34195953.ece>
 32. BigBasket Faces Data Breach: Details Of 2 Cr Users Put On Dark Web, Inc42, November 2020, <https://inc42.com/buzz/big-breach-at-bigbasket-details-of-2-cr-users-put-on-dark-web-claims-cyble/>
 33. Nikhil Subramaniam, Flipkart Gets Caught In BigBasket Data Leak Aftermath As Users Allege Attacks, May 2021, <https://inc42.com/buzz/flipkart-gets-caught-in-bigbasket-data-leak-aftermath-as-users-allege-attacks/>
 34. 'Guidelines For MNEs - Organisation for Economic Co-Operation and Development' (Mne Guidelines.oecd.org) <http://mneguidelines.oecd.org/>
 35. David Wright, Rowena Rodrigues and David Barnard-Wills, 'AI And Cybersecurity: How Can Ethics Guide Artificial Intelligence To Make Our Societies Safer? - Trilateral Research' (Trilateral Research) <https://www.trilateralresearch.com/ai-and-cybersecurity-how-can-ethics-guide-artificial-intelligence-to-make-our-societies-safer/>
 36. 'ISO 27003: Guidance For ISO 27001 Implementation' (ISMS.online) <https://www.isms.online/iso-27003/>
 37. 'Principle On Robustness, Security And Safety (OECD AI Principle) - OECD.AI' (Oecd.ai, 2021) <https://oecd.ai/dashboards/ai-principles/P8>
 38. 'Openai Charter' (OpenAI, 2018) <https://openai.com/charter/>
 39. Karen Hao, 'The Messy, Secretive Reality behind Openai's Bid To Save The World' (MIT Technology Review, 2020) <https://www.technologyreview.com/s/615181/ai-openai-moonshot-elon-musk-sam-altman-greg-brockman-messy-secretive-reality/>

- 
- 
40. 'Justice K. S. Puttaswamy v. Union Of India' (Scobserver.in, 2012) <https://www.scobserver.in/court-case/fundamental-right-to-privacy>
 41. Sheenu Sura, 'Understanding The Concept Of Right To Privacy In Indian Perspective - Ignited Minds Journals' (Ignited.in, 2019) <http://ignited.in/I/a/89517>
 42. Asia J. Biega and others, 'Operationalizing the Legal Principle of Data Minimization for Personalization' (Arxiv.org, 2020) <https://arxiv.org/pdf/2005.13718.pdf>
 43. 'Predictive Model for Tax Return Checks' (Skatteetaten.no, 2016) <https://www.skatteetaten.no/globalassets/om-skatteetaten/analyse-og-rapporter/analysenytt/analysenytt2016-1.pdf#page=12>
 44. 'National Strategy for Artificial Intelligence' (Niti.gov.in, 2018) <https://niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf#page=85>
 45. 'Inclusive Growth, Sustainable Development And Well-Being (OECD AI Principle) - OECD.AI' (Oecd.ai) <https://oecd.ai/dashboards/ai-principles/P5>
 46. 'Ethics Guidelines for Trustworthy AI | Shaping Europe's Digital Future' (Ec.europa.eu) <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>
 47. (Pdpc.gov.sg) <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf>
 48. 'Ethics, Transparency And Accountability Framework For Automated Decision-Making' (gov.uk, 2021) <https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making>
 49. 'Data Ethics Framework' (gov.uk, 2020) <https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020>
 50. 'Explainable AI: The Basics' (Royalsociety.org, 2019) <https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf>
 51. 'AI Principles - Future Of Life Institute' (Future of Life Institute, 2017) <https://futureoflife.org/ai-principles/>
 52. 'Montreal Declaration For A Responsible Development Of Artificial Intelligence - La Recherche - Université De Montréal' (Recherche.umontreal.ca, 2017) <https://recherche.umontreal.ca/english/strategic-initiatives/montreal-declaration-for-a-responsible-ai/>
 53. 'IEEE SA - The IEEE Global Initiative on Ethics Of Autonomous And Intelligent Systems' (Standards.ieee.org, 2021) <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>
 54. 'Tenets - The Partnership On AI' (The Partnership on AI, 2021) <https://www.partnershiponai.org/tenets/>
 55. 'AI in the UK: Ready, Willing and Able?' (Publications.parliament.uk, 2019) <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf#page=417>
 56. 'Research And Innovation' (European Commission - European Commission, 2018) https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf
 57. 'The Toronto Declaration: Protecting The Right to Equality and Non-Discrimination in Machine Learning Systems' (Accessnow.org, 2018) https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf
 58. Nicholas Diakopoulos and others, 'Principles For Accountable Algorithms And A Social Impact Statement For Algorithms: FAT ML' (Fatml.org) <https://www.fatml.org/resources/principles-for-accountable-algorithms>
 59. 'Text - H.R.2231 - 116Th Congress (2019-2020): Algorithmic Accountability Act Of 2019' (Congress.gov, 2019) <https://www.congress.gov/bill/116th-congress/house-bill/2231/text>
 60. Dillon Reisman and others, 'Algorithmic Impact Assessments: A Practical Framework For Public Agency Accountability' (Aino Institute.org, 2018) <https://ainowinstitute.org/aiareport2018.pdf>

- 
- 
61. 'The Struggles Businesses Face in Accessing the Information They Need' (Info.pure storage.com)
http://info.purestorage.com/rs/225-USM-292/images/Big%20Data%27s%20Big%20Failure_UK%281%29.pdf?aliId=64921319
 62. 'Feature Selection - Wikipedia' (En.wikipedia.org) https://en.wikipedia.org/wiki/Feature_selection
 63. Saikat Guha, Bin Cheng and Paul Francis, 'Privad: Practical Privacy In Online Advertising' (Usenix.org)
https://www.usenix.org/legacy/event/nsdi11/tech/full_papers/Guha.pdf
 64. Leyla Kazemi and Cyrus Shahabi, 'TAPAS: Trustworthy Privacy-Aware Participatory Sensing' (Infolab.usc.edu, 2012)
<https://infolab.usc.edu/DocsDemos/kazemi-TAPAS-KAIS.pdf>
 65. Rachel Levy Sarfin, '5 Characteristics of Data Quality - See Why Each Matters to Your Business' (Precisely, 2021)
<https://www.precisely.com/blog/data-quality/5-characteristics-of-data-quality#:~:text=There%20are%20data%20quality%20characteristics,read%20on%20to%20learn%20more.>
 66. 'REPORT On a Comprehensive European Industrial Policy on Artificial Intelligence and Robotics' (Europarl.europa.eu, 2019) https://www.europarl.europa.eu/doceo/document/A-8-2019-0019_EN.html#title4
 67. Thomas C. Redman, 'Bad Data Costs The U.S. \$3 Trillion Per Year' (Harvard Business Review, 2016)
<https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>
 68. Angus Loten, 'AI Efforts At Large Companies May Be Hindered by Poor Quality Data' (WSJ, 2019)
<https://www.wsj.com/articles/ai-efforts-at-large-companies-may-be-hindered-by-poor-quality-data-11551741634>
 69. Sylvia Kingsmill and Ann Cavoukian, 'Privacy by Design' (www2.deloitte.com)
<https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/risk/ca-en-ers-privacy-by-design-brochure.PDF>
 70. Ann Cavoukian, 'Privacy by Design: The 7 Founding Principles' (Ipc.on.ca, 2011)
<https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf>
 71. 'Software Development with Data Protection by Design and by Default' (Datatilsynet, 2017)
<https://www.datatilsynet.no/en/about-privacy/virkosomhetenes-plikter/innebygd-personvern/data-protection-by-design-and-by-default/>
 72. Ann Cavoukian, 'Operationalizing Privacy by Design: A Guide To Implementing Strong Privacy Practices' (Collections.ola.org, 2012) <https://collections.ola.org/mon/26012/320221.pdf>
 73. Ian J. Goodfellow and others, 'Generative Adversarial Nets' (Papers.nips.cc)
<https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>
 74. 'Federated Learning: Collaborative Machine Learning without Centralized Training Data' (Google AI Blog, 2017)
<https://research.googleblog.com/2017/04/federated-learning-collaborative.html>
 75. Geoffrey Hinton, Sara Hinton and Nicholas Frosst, 'Matrix Capsules With EM Routing' (Openreview.net, 2018)
<https://openreview.net/pdf?id=HJWlFGWRb>
 76. Zhanglong Ji, Zachary C. Lipton and Charles Elkan, 'Differential Privacy And Machine Learning: A Survey And Review' (arxiv.org, 2014) <https://arxiv.org/abs/1412.7584>
 77. 'Homomorphic Encryption Standardization – An Open Industry / Government / Academic Consortium to Advance Secure Computation' (Homomorphic Encryption.org) <http://homomorphicencryption.org/>
 78. 'Machine Learning Research Group | University Of Texas' (cs.utexas.edu)
http://www.cs.utexas.edu/~ml/publications/area/125/transfer_learning
 79. 'RAIRD' (Raird.no, 2019) <http://raird.no/>

Deutsche Gesellschaft für
Internationale Zusammenarbeit
(GIZ) GmbH

A2/18 Safdarjung Enclave
New Delhi-110029 India

T: +91-11-49495353

E: nrm@giz.de

www.giz.de/India

