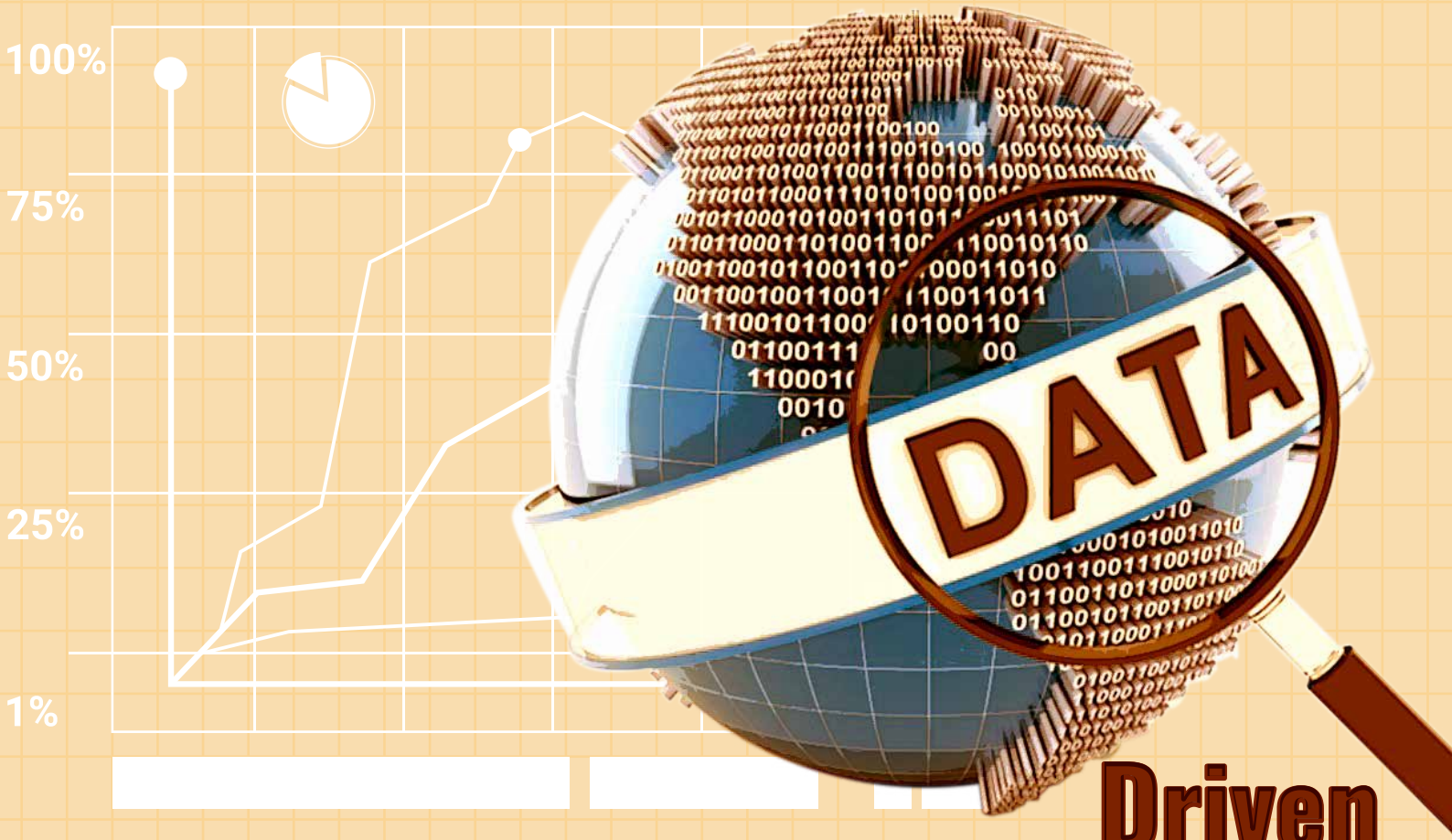
VISIONIAS
INSPIRING INNOVATION

Weekly Focus

#26 Week 5, Jan 2021



Data Driven Innovations and Privacy

Introduction

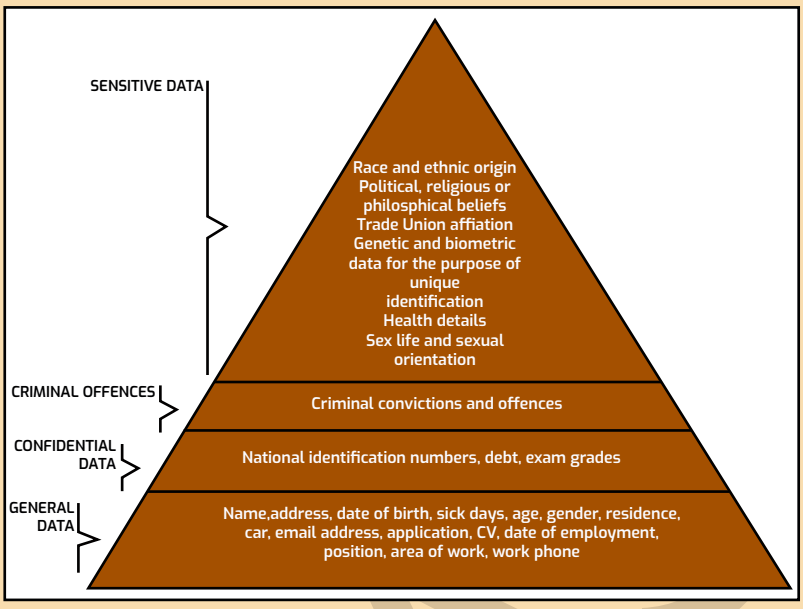
Data-driven innovation forms a key pillar of the 21st century sources of growth. The confluence of several trends, including increasing migration of socio-economic activities to the digital space and the decline in the cost of data collection, storage and processing are leading to generation and use of huge volumes of data – commonly referred to as "big data". These large data sets are becoming a core asset to the economy, fostering new industries, processes and products and creating significant competitive advantages. But the ability of public and private entities to collect and process detailed and potentially intrusive data about people, both easily and cheaply, has the potential to generate substantial privacy problems and vulnerabilities, thus bringing the idea of data regulation and data protection to the fore.

Governments worldwide are striving to establish a proper regulatory response to the ongoing practices of personal-data collection, analysis, and usage. However, the presence and content of such regulations may directly or indirectly influence the extent and direction of data-based innovation. In this scenario, it is imperative to first understand what is data, its types and data privacy? How is personal data driving innovation in the present times? What are the implications of data driven innovation on the privacy of users? What are the consequences of privacy regulation for innovation outcomes? And how can we promote data privacy whilst protecting innovation? In this edition, we will attempt to answer these questions.

What is data, what are its types and what is data privacy?

Data means and includes a representation of information, facts, concepts, opinions, or instructions in a manner suitable for communication, interpretation, or processing by humans or by automated means. In terms of presence/absence of identifiable information related to consumers, data can be broadly **classified into two types:**

- **Personal data:** It includes data relating to a natural person (generally referred to as data principal) that directly or indirectly identifies that individual. Such data may include information regarding any characteristic, trait, attribute or any other feature of the identity of such natural person. Personal data can be further categorized into-
  - **Sensitive Personal Data:** This may include sensitive information whose misuse can cause substantial harm to an individual, such as passwords; financial data; health data; official identifier; sexual orientation; biometric data; genetic data; caste or tribe; religious or political belief etc.
  - **General/ordinary personal data:** Ordinary personal data may include personal identification details such as name and address, customer relationships, tax-related matters etc.



  - Other forms of personal data include **information relating to criminal convictions and offences** among others.
- **Non-personal data:** It includes data that is either completely unrelated to an identifiable natural person; or which was initially personal but was later aggregated and made anonymous. There are broadly three types of non-personal data:
  - **Public non-personal data:** owned by governments, such as anonymized land records and vehicle registration data.
  - **Private non-personal data:** owned by non-government players and derived from assets or processes that are privately-owned.
  - **Community data:** which is the raw data of a group of people that may also be collected by private players. This may include some data collected by the municipal corporations, public electric utilities, and private players like telecom, e-commerce, and ride-hailing companies.

Generation of large scale of data especially in the form of personal data brings into picture the idea of data privacy. **Data Privacy** or Information privacy encompasses **3 key elements:**

- Right of an individual to be left alone and have **control over their personal data.**
- **Procedures for** proper handling, processing, collecting, and sharing of **personal data.**
- Compliance with **data protection laws.**

How is personal data driving innovation in the present times?

Advance developments in the field of Artificial Intelligence, Machine Learning, Big data analytics, Cloud computing etc. has changed the extent and nature of how information is collected, stored and analyzed in today's digitized world. Some prominent areas where personal data is currently driving innovation are as follows:

- **Online advertising/ Targeted ads:** Through collection and use of large amounts of customer data, the advertising industry can now determine which kind of customers are most likely to be influenced by a particular ad and whether the advertising has actually succeeded. For example, search engines like Google place ads on search result pages based on one's browsing histories.

- **Health care:** Collection and analysis of large data sets of patient's electronic medical records (EMRs) or health records can assist health care providers to make diagnostic predictions and treatment suggestions. For example, using devices like Fitbit, doctors can help understand Blood Pressure patterns of a patient enabling a more precise intervention.
  - Also, such records can be used by research bodies and public health institutions to facilitate research and development in healthcare solutions and policy formulation.
- **Operational Efficiency:** Wide-scale collection of consumer data can also enhance a firm's operational efficiency through-
  - **Effective Management of supply chains.**
  - **Development of Recommender systems** to offer recommendations according to user's preferences, interest, or observed behavior about any product.
  - **Development of customised products** as per consumers' needs and interests.
- **Governance:** Personal data can be used by state machinery for purposes such as targeted delivery of social welfare benefits, effective planning and implementation of government schemes, etc. For example, the idea of **Data based policing** can enable more efficient utilization of available human resources.
- **Other potential innovative outcomes:**
  - **Finance and Insurance:** Data represents a unique opportunity for most banking and financial services organizations to leverage their customer data to transform their business, realize new revenue opportunities, manage risk, and address customer loyalty.
  - **Transport:** An individual's personal locational data could be used for monitoring traffic and improving driving conditions on the road.

Impact of COVID-19 on Data Privacy and Innovation

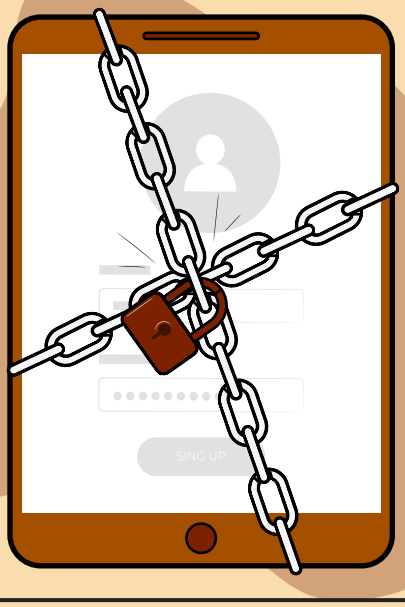
Developments:

- **Rapid deployment of innovative digital technologies** such as contact-tracing apps and facial recognition.
- **Intensification and rushed adoption in the use of pre-existing digital products,** such as video-conferencing software.
- **Increased use of digital surveillance measures by many private companies** to track the activities of their employees that are working from home.

Concerns generated from these developments:

- **Surveillance creep:** The pandemic could be exploited as an opportunity to normalize governmental surveillance particularly in the domestic and biopolitical sphere.
- **Opaque data processing:** Governments and health authorities have not provided enough transparency about how the tracking works.
- **Function creep** (information being used for a purpose that is not the originally specified purpose): There have been concerns about potential transfer of data from telecommunication companies to law enforcement agencies.
- **Regulatory vacuum:** The pandemic has highlighted the absence of data protection framework specifically designed for use in emergency situations.
- **Involuntary consent:** Rushed and sometimes forced implementation and use of video-conferencing tools or contact tracing applications raises the question of whether consent to the use of rushed innovations can still be considered voluntary during a pandemic.
- **Ad Hoc deployment:** Digital technologies have been deployed on an ad-hoc basis without proper impact assessment, stakeholder consultation or evaluation, which raises the possibility of data breaches and other privacy concerns.

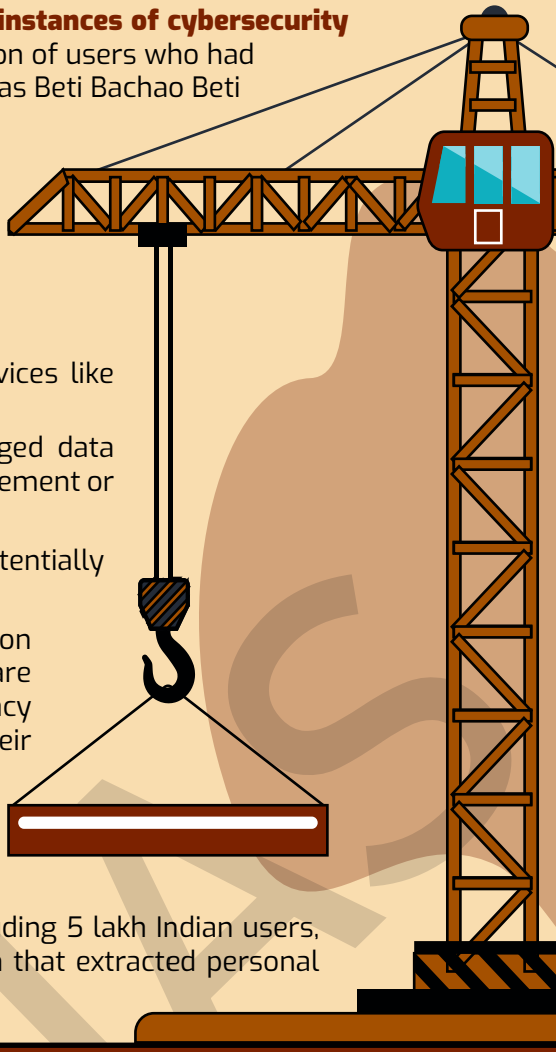
What are the implications of data driven innovation on the privacy of users?



While data can be put to beneficial use, the **unregulated and arbitrary use of data,** especially personal data, has the potential to threaten the privacy and autonomy of an individual. Some concerns in this regard are-

- **Data breaches:** Data breaches can take many forms including hackers gaining access to data through a malicious attack; lost, stolen, or temporary misplaced equipment; employee negligence; and policy and/or system failure.

- Depending on the scale and nature, data breaches and disclosure of personal data in the public domain can have adverse effects on privacy of users including **public humiliation, identity theft, financial fraud, harassment, stalking etc.**
- Government data shows that in 2019 alone, **India witnessed 3.94 lakh instances of cybersecurity breaches.** For instance, recently databases including personal information of users who had donated funds to the PM Relief Fund and several other such funds such as Beti Bachao Beti Padhao were leaked on the dark web.



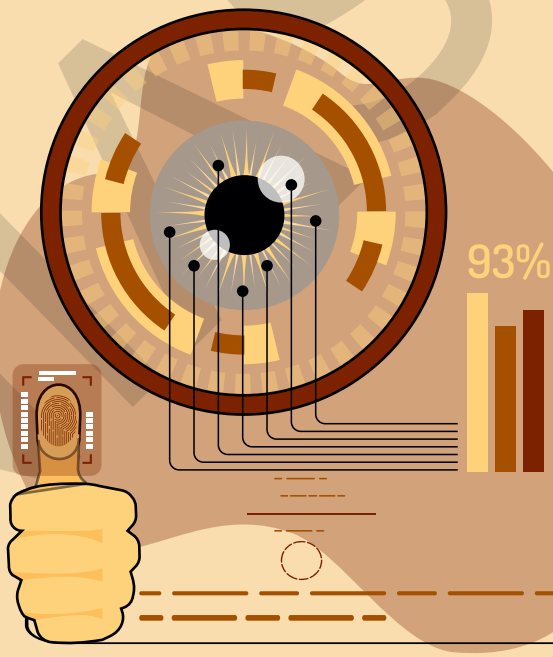
- **Profiling of individuals:** Behavioral profiling of a person through big data analysis can be used to predict their interests and behavior. Derivations based on such profiling can be **used for unequal treatment or discrimination.** For example, profiling could be used by politically motivated organizations for targeting and discriminating against people with a certain political inclination.
  - Also, decisions based on profiling can lead to refusal of financial services like insurance or a credit card.
- **Erosion of individual autonomy:** Continuous data collection and prolonged data storage can place direct or indirect restriction on an individual's speech, movement or any other action arising out of a fear of being observed or surveilled.
  - Also, data on user preferences and activities on online platforms can potentially be used to influence opinions, jeopardizing **autonomous thinking.**
- **Lack of real alternatives:** In current times, the field of data driven innovation is dominated by large search engines and social networking firms. Users are usually left with no real alternatives even if they disagree with the privacy policies of these firms. This restricts the choices of users to safeguard their own privacy.
- **Potential for data misuse:** Data gathering practices are usually opaque, mired in complex privacy forms that are unintelligible, thus leading to data usage and practices that users have little control over.
  - For instance, Facebook admitted that the data of 87 million users, including 5 lakh Indian users, was shared with Cambridge Analytica through a third-party application that extracted personal data of Facebook users who had downloaded the application.

Existing Approaches to Data Privacy and Protection

- **International**
  - **European Union:** EU's **General Data Protection Regulation (GDPR)** is considered to be one of the most stringent data protection laws in the world and follows a **rights based approach ,thus placing the** individual at the centre of the law.
    - It imposes extensive control over the processing of personal data both at the time of, and after the data has been collected and collection of sensitive personal data is prohibited subject to certain exceptions.
    - The GDPR also imposes fines for non-compliance with maximum fine going up to €20 Million or 4% of annual worldwide turnover – whichever is greater.
  - **United States:** The US follows a **laissez faire approach** and does not have an overarching data protection framework. The US approach to data protection thus has two discernible trends— stringent norms for government processing of personal information; and notice and choice based models for private sector data processing. There is no comprehensive set of privacy rights/principles that collectively address the use, collection and disclosure of data in the US. Instead, **there is limited sector specific regulation.**
  - **China:** In terms of data protection, China primarily focuses on **data localization** under the Chinese Cybersecurity Law which states that Chinese citizen's personal information and important data, which are collected and generated by critical information infrastructure (CII) operators in China must be stored domestically on Chinese servers.
- **Domestic**

- **Personal Data Protection Bill, 2019:** India is in the process of creating its first **cross-sectoral legal framework for data protection** via this law. The Bill derives its inspiration from a previous draft version prepared by a committee headed by retired **Justice B N Srikrishna.**
  - The bill proposes requirements that all entities processing data will have to comply with. These include **consent framework, privacy by design, limitation on purpose and data storage and collection, grounds for processing of personal data and sensitive personal data, transparency requirements,** localization requirements, and the creation of grievance-redress systems etc.
- **Data Privacy as a Fundamental Right:** Supreme Court in **K.S Puttuswamy v Union of India** case, recognised right to privacy as a fundamental right emerging primarily from right to life and personal liberty under Article 21 of the Constitution. The judgment also declared **informational privacy to be a subset of the right to privacy.**
- **Draft Health Data Management Policy:** It aims to safeguard digital personal data within the ambit of the National Digital Health Ecosystem (NDHE), including the Personal Health Identifier, the electronic health records and electronic medical records, by implementing adequate technical and organizational measures across the NDHE.
- **Information Technology Act, 2000:** The Information Technology (Reasonable Security Practices and Procedures and Sensitive Personal Data or Information) Rules, 2011 (SPD Rules) were issued under **Section 43A** of the act which holds a body corporate liable for compensation for any negligence in implementing and maintaining reasonable security practices and procedures while dealing with sensitive personal data or information.
- **Financial Data:** Data related to the Banking Sector is regulated under Credit Information Companies (Regulation) Act, 2005, Credit Information Companies Regulations, 2006, circulars of Reserve Bank of India (RBI) including KYC circulars etc.
  - RBI in April 2018 also put out a circular requiring that all data relating to payment systems should be stored in a system only in India.

What are the possible consequences of privacy regulations for innovation outcomes?



The nature, structure and processes of the privacy regulation framework can have positive as well as negative consequences for innovation outcomes.

Positive consequences:

- **Enabling effective and high-quality data collection:** Assurance provided by privacy regulations to consumers, that their data will not be used for malicious purposes, will **encourage voluntary sharing of data,** enhancing the quality and quantity of data collection.
  - Also, prevalence of clear regulations creates the possibility of **active collection of data** for the sole purpose of innovation.
- **Creating a niche market:** Privacy rules will call for the development of new and innovative systems and products, commonly referred to as **Privacy Enhancing Technologies,** for respecting or maintaining privacy on the market.
  - Also, development of such technologies will indirectly **decrease the cost of ensuring data privacy.** For example, creation of more robust and economically cheaper encryption algorithms.
- **Generating demand through trust:** Commercial conduct that respects consumer privacy (especially in the context of e-commerce and other long-distance transactions) is crucial for generating trust. This trust will generate consumer demand, which in turn will drive innovation.
- **Facilitating creativity:** Data privacy regulations create a sense of security against issues like surveillance and profiling. This encourages individuals to think freely and thus enables innovation.
- **Enhancing competition:** Privacy laws will require large firms to cut back on certain data analysis and usage practices. Thus, new competitors will be able to enter these markets and offer competitive services as there will be a level playing field vis-à-vis access to data. Both incumbents and competitors will innovate as part of this competitive dynamic.

Negative consequences:

- **Escalation of costs:** Maintaining full compliance with restrictive privacy laws can be costly, particularly since adherence can result in a loss of valuable marketing data. This can indirectly discourage entrepreneurship, investment and innovation.
  - For instance, a study estimated that domestic investment in India could see a reduction of up to 1.4 percent due to localization requirements.
- **Business Uncertainty:** As the new technologies continuously evolve, the privacy laws also keep changing. This creates uncertain regulatory environment for innovation.
- **Impacts of Precautionary Privacy Regime:** If the public policy is primarily guided by the precautionary principle (that emphasizes caution, pausing and review before leaping into new innovations), it may discourage entrepreneurial ideas and the start up culture.
- **Creating oligopolistic market structures:** Stringent privacy regulations may disproportionately damage small companies, start-ups etc. and may reduce competition, due to reasons such as-
  - **Challenges in complying with regulations:** Technical and financial capabilities required for fulfilling obligations under regulations raises entry barriers for smaller firms, thus reducing competition and incentives for innovation. For example, compliance cost of regulations and inconsistency with regulations may trigger altogether product abandonment from several start-ups.
  - **Creation of walled gardens on the Internet:** Walled gardens are where users could be confident that the websites visited were safe in terms of both computer security and reliability of content. Difficulty in understanding privacy notices could give large firms an advantage over small firms in terms of consumer trust, leading users to spend an increasing portion of their online time within the walled-garden environments provided by large firms.
  - **Data monopoly:** Strong privacy rules and measures might render access to online users extremely difficult to any entity except those already controlling the digital infrastructure. This will indirectly create data haves and have-nots in the innovation ecosystem which will negatively affect new market entrants especially data-driven start-ups.
- **Hindered information flow:** Privacy laws can substantially impede the flow of personal data as a result of which innovators are unable to use these data flows optimally to produce novel products and services.
  - Also, disclosing criteria increases public scrutiny, resulting in potential limitations in data sharing.

Case study of GDPR: The General Data Protection Regulation (GDPR) went into effect in 2018. Now, there is mounting evidence that the law has not produced its intended outcomes, highlighted by concerning trends such as:

- **Reduced Competition in Digital Advertising:** Advertising vendors have lost market reach in the EU, particularly smaller players—who lost between 18 and 31 percent between April and July 2018.
- **Failure to Increase Trust:** 81 percent Europeans who provide personal information online feel they have no control or partial control over this information.
- **Complicated for Consumers to Understand:** 63 percent Europeans have never heard of the GDPR or do not know exactly what it is.
- **Strains Resources of Regulators:** Companies reported spending an average of $1.3 million in 2017 on GDPR compliance and were expected to spend an additional $1.8 million in 2018.

How to promote data privacy while protecting innovation?

The challenge is to design an adaptive policy environment that addresses exploitation and deception while leaving ample space for innovation. Several steps can be taken in this direction:

- **Anonymization of personal data sets:** Personal data can be turned into information that cannot be used to identify individuals through techniques like-
  - **Pseudonymization/ De-identification:** It masks personal data by replacing identifying information with unique artificial identifiers.
  - **Anonymization:** It is an irreversible process of transforming or converting personal data to a form in which a data principal cannot be identified through any means reasonably likely.

- **Transforming information into statistical data:** Firms can avoid collecting personal data and can perform data analytics on top of such statistical data.

- **Collecting less data and making optimal use of it:** Companies can increase customer satisfaction and improve long-term customer loyalty by collecting only the data that is actually relevant.
  - In the long term, firms need to use the data they have collected more efficiently, store less data, and ensure that data collected in this way is given optimal protection.
- **Revamping regulatory regimes:** Preventive regulatory obligations should be layered, based on an assessment of their costs and benefits. Obligations for firms that do not process data intensively or that do not handle sensitive personal data should be reduced in a manner commensurate to the risks from their activities.
  - Also, regulatory uncertainty must be reduced by minimizing use of ambiguous language in the laws.
- **Promoting technological innovations:** In particular, technology can provide the means for privacy-respecting innovations to take place. Some examples of innovations driven by privacy requirements:
  - **Synthetic data generation** using machine learning systems to help data scientists access personal data anonymously.
  - Identity management fueled by **Blockchain techniques which allow to verify identities without exposing personal data.**
  - Personal data monetization by allowing individuals to get value from sharing their data.
  - Consent dashboards and **user-controlled data vaults.**
- **Encouraging privacy by design:** The principle establishes data handling practices in the organization in a manner ensuring compliance with the law by minimizing or eliminating adverse impacts on privacy.
- **Creating universally-available data sets:** Enhanced access to this data would level the playing field and allow start-ups to grow their businesses according to the quality of their technical and business innovations, potentially allowing them to compete with the current tech giants.
  - An accredited body or regulator can be constituted with the power to sign off on companies' data policies and to help make anonymous, fair and balanced open data sets publicly available.
- **Data empowerment:** Data empowerment is the process where people, on their own or with the help of intermediaries, take control or gain the power to take control of their data to promote their and their society's wellbeing. Improving people's understanding of general data standards and good data practices can improve behaviors and outcomes.

Conclusion

Technological gains should not come at the expense of privacy. Privacy is a fundamental and inalienable right. Regulation should not try to balance innovation and privacy, but instead promote appropriate innovation that is based on respect for privacy and user control. As long as we create regulation that is predictable, nimble and fair, there is no reason to believe that regulation will impede innovation. In fact, trust will create the right environment for greater participation in the digital economy, thus strengthening businesses and innovators. Therefore, innovation and privacy can be made compatible by striking a balance between effective data gathering and a respect for consumer choice in matters of privacy.

TOPIC AT A GLANCE

Data and Data Privacy

Data: Representation of information, facts, concepts, opinions, or instructions in a manner suitable for communication, interpretation, or processing by humans or by automated means.

Personal data: Data relating to a **natural person** that directly or indirectly identifies that individual. It includes **Sensitive Personal Data,** such as passwords; financial data; health data; sexual orientation etc.; **General/ordinary personal data** such as name and address, customer relationships and other forms of personal data include information relating to criminal convictions and offences.

Non-personal data: Data that is either completely unrelated to an identifiable natural person; or which was initially personal but was later aggregated and made anonymous.

Data Privacy or Information privacy encompasses **3 key elements:**
- Right of an individual to be left alone and have **control over their personal data.**
- **Procedures for** proper handling, processing, collecting, and sharing of **personal data.**
- Compliance with **data protection laws**

Fields in which Personal data is presently driving innovation
- Online advertising/Targeted ads, Health care, Governance, Finance and Insurance, Transport.
- Operational Efficiency - Effective Management of supply chains, Development of Recommender systems, Development of customised products.

Implications of data driven innovation on the privacy of users
- **Data breaches:** Can have adverse effects on privacy of users including public humiliation, identity theft, financial fraud, harassment, stalking etc.
- **Profiling of individuals:** Can be used for unequal treatment or discrimination.
- **Erosion of individual autonomy:** Can place direct or indirect restriction on an individual's speech, movement and can influence opinions, jeopardizing **autonomous thinking.**
- **Lack of real alternatives:** Restricts the choices of users to safeguard their own privacy.
- **Potential for data misuse:** Users have little control over data usage and practices.

Consequences of privacy regulations for innovation outcomes

Positive consequences
- **Enabling effective and high-quality data collection:** by encouraging voluntary sharing of data.
- **Creating a niche market:** in the field of Privacy Enhancing Technologies.
- **Generating consumer demand through trust.**
- **Facilitating creativity:** by creating a sense of security against issues like surveillance and profiling.
- **Enhancing competition:** by establishing a level playing field vis-à-vis access to data.

Negative consequences
- **Escalation of costs:** to maintain full compliance with restrictive privacy regulations.
- **Business Uncertainty:** due uncertain regulatory environment surrounding technology and privacy laws.
- **Precautionary principle** (that emphasizes caution, pausing and review before leaping into new innovations) may discourage entrepreneurial ideas and start up culture.
- **Creation of oligopolistic market structures:** Disproportionate damage to small companies, start-ups etc. due to reasons such as **technical and financial challenges in complying with regulations, creation of walled gardens on the Internet and data monopoly.**
- **Hindered information flow.**

Way Forward: Promoting data privacy while protecting innovation
- **Anonymization of personal data sets:** through techniques like Pseudonymization/ De-identification, Anonymization and Transforming information into statistical data.
- **Collecting only relevant data and making optimal use of it.**
- **Revamping regulatory regimes.**
- **Promoting privacy-respecting technological innovations.**
- **Encouraging privacy by design** (data handling practices that ensures compliance with the law by minimizing or eliminating adverse impacts on privacy).
- **Creating universally-available datasets** to enhance access to data for startups.
- **Data empowerment:** Process where people, on their own or with the help of intermediaries, take control or gain the power to take control of their data.