https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

Review Article ISSN: 3049-3072

IJAS-25-128

Big Data and Great Privacy Challenges in the Digital Era - A Comprehensive Study

Muhammad Rawish Siddiqui*

Department of Cybersecurity & Data Privacy, Paris American International University (PAIU), Paris, France

Corresponding Author: Muhammad Rawish Siddiqui, Department of Cybersecurity & Data Privacy, Paris American International University (PAIU), Paris, France

Received date: 09 December, 2025, **Accepted date:** 23 December, 2025, **Published date:** 30 December, 2025

Citation: Siddiqui MR (2025) Big Data and Great Privacy Challenges in the Digital Era - A Comprehensive Study. Innov J Appl Sci 2(6): 42.

Abstract

In the era of globalization, industries are undergoing a rapid digital transformation, driven by big data ecosystems, conducting data intensive operations and interconnected technologies, generating huge data by means of IoT devices, AI systems, social media, Edge processing, smart systems and autonomous machines, all are designed to collect, process and transmit multi-modal data continuously and at a high rate. Even though these innovations enhance overall efficacy, productivity and competitiveness, they also present significant unavoidable data privacy challenges that threaten data breaches, operational continuity and regulatory compliance.

Bearing in mind the criticality of the topic, this study targets to identify and evaluate the most critical big data privacy challenges, analyses their probability of occurrence, contextual relevance to the aviation and consequential impact on industrial operations. Although this research is industry-agnostic; however, illustrative examples were drawn from the aviation industry perspective to demonstrate applicability.

A combined approach of systematic literature review and risk assessment frameworks, is used to examine mention challenges. The research highlights prominent challenges related to data privacy within the big data ecosystem, that can lead to multiple severe consequences like brand damage, loss of trust, business disruptions, financial losses and other undesirable outcomes. Finally, this study presents actionable workarounds to overcome existing challenges in the domain of big data privacy, followed by potential directions for further research.

Keywords: Big data, Data privacy, Personal data protection, Data security, Data regulation, Data compliance, GDPR, PDPL, CCPA, Data privacy challenges.

Introduction

As stated by Demirol, the concept of data is present in every era of human existence. In different periods of time, data is stored by using different methods and data storage techniques are continuously developing today. In today's digital era, massive amounts of data are being generated every second, from financial statements, social media posts, mobile apps, online shopping, smart devices, security cameras, IoT (Internet of Things) devices and much more. This exponential growth, massive amount and variety of data is collectively referred to Big Data. In other words, Big Data in its nature, is large (Volume), fast (Velocity) and complex (Variety). Big Data helps organizations to leverage and infer insight to make better decisions, improve services, enhance quality, even predict prospective trends and design future business roadmaps. However, this incredible power and remarkable benefits entail severe and unavoidable privacy and security challenges. Sensitive and valuable data may be leaked for threat, blackmail and financial gain by attackers. Therefore, protecting Big Data appropriately and ensuring privacy throughout its lifecycle is crucial [1] (Table 1).

Data sources	Ingestion	Storage	Processing	Analytics	Applications
Web (Websites, Logs)	Batch ETL (Sqoop, Airflow, dot)	Data Lake (HDFS, S3, ADLS)	Batch Processing (Hadoop MapReduce, Spark)	ML/A (Training, MLOps, Model Registry)	Fraud Detection (Anomaly & Risk Scoring)
Mobile (Apps, SDKs)	Streaming (Kafka, Fink, Spark Streaming)	Object Storage (S3, GCS, Azure Blob)	Stream Processing (Flink, Storm, Spark Structured Streaming)	BI Dashboards (Looker, Power BI, Tableau)	Personalization (Recommendations, 360° Profiles)
IoT & Sensors (Edge, Telemetry)	API Gateways (REST, RPC, Webhooks)	Data Warehouse (Snowflake, BigQuery, Redshift)	Real-time Serving (Feature Stores, Low-latency Views)	Predictive & Prescriptive (Forecasting, Optimization)	Reports & BI (Operational & Executive)
Social media (Twitter/X, FB, Insta)	Change Data Capture (Debezium, GoldenGate)	NoSQL (Cassandra, MongoDB, DynamoDB)	Data Marts & Curated Zones (Star/Snowflake Schemas)	Visualization Tools (Notebooks, Charts, GeoMaps)	Al-powered Apps (Copilots, Chat, Agents)

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

APIs (Partner, Public)	Data Quality & Governance (Validation, Pll Masking)	RDBMS (PostgreSQL, MySQL, Oracle)	Metadata Catalog (Data Lineage, Glossary)	Data Science Workbench (Jupyter, Spark Notebooks)	APIs/Exports (Data Sharing & Monetization)
Databases (OLTP, ERP, CRM)	Orchestration (Airflow, Dagster, Prefect)	Blockchain (Immutable Ledger/Audit)	Data Transformation (ETL/ELT)	Operation Analytics	Business Applications
Cloud Services (SaaS, SaaP Events)	Event Streaming/Webhooks	Cloud storage/ Data Lake	Event processing	Usage analytics	Automation and integration
Other Logs (Clickstream, Server Logs)	Log collectors (Fluentd, Logstash)	Log storage/Data Lake	Log processing and aggregation	Behavioral analytics	Monitoring and observability

Table 1: Core data pipeline components.

Research objectives

Indeed, organizations collect, maintain and analyze growing volumes of data, from personal identifiers (names, age, sex, mobile, location etc.) to behavioral trends, IoT sensor logs and confidential organizational information such as organization hierarchy, internal taxonomies and ontologies, as a result, increasing the complexity from Big Data Security and Privacy perspectives. As shown in Table 1, Big Data systems differ from traditional databases as Big Data doesn't just involve one source or one system, instead it draws from multi-modal sources, in distinct formats and at very high speed. "All of this information comes from a variety of places, including social media, sensors, scientific applications, surveillance, video and image archives, Internet search indexing, medical records, corporate transactions and system logs" and eventually, making it difficult to monitor, control, or fully understand the whole data lifecycle, how data is being collected, stored, used, shared, or even archived [2]. Often, data includes meaningful insights that are not directly linked to an individual but can still infer sensitive patterns about groups, processes, or organizational operations. Protecting privacy in this case, requires an appropriate and broader approach that addresses all privacy related dimensions and domains (i.e., Legal, Technical, Political and Ethical) mentioned by Sajid Momin et al. and Zhou et al., respectively [3,4]. Through this research, we aim to help researchers, practitioners and decision-makers to understand:

- Big data privacy dimensions, major components and layers.
- What are the key privacy challenges that exist at different stages of the Big Data Lifecycle, Occurrence Probability (on a 0-1 scale) and how do they impact individuals and organizations?
- How do emerging technologies like IOT, AI, LLM, Blockchain, Biometrics, Surveillance, Wearables Devices and Edge Computing introduce new privacy challenges and risk in Big Data ecosystem?
- How does each identified challenge relate to the operations and sustainability of the Aviation Industry?
- Despite the availability of existing solutions, unresolved challenges continue to hinder their effectiveness. What innovative framework could be formulated for further academic or practical exploration?

Another important part of this study is the legal and regulatory angle. Laws and regulations like GDPR (General Data Protection Regulation) in Europe, PDPL (Personal Data Protection Law) in KSA and CCPA (California Consumer Privacy Act) in the U.S. have introduced strong rules for data privacy. But application of these laws and regulations to Big Data, is not always easy, "the scope of work a privacy team assumes is large, given the variety and sheer number of global, federal and local privacy, or privacy-impacting, laws" [5].

The primary focus of this study is on big data privacy challenges. However, study also aims to contextualize each major challenge within the aviation industry, assess its likelihood of occurrence and analyze its potential consequences.

In conclusion, this study not only consolidates existing knowledge from past research but also points out some critical gaps, where more research is crucial. These include improving technologies, making privacy laws more practical for Big Data and developing new methodologies to protect privacy without stopping innovation. By presenting a comprehensive overview of current risks, potential solutions and unresolved challenges, finally, this study introduces a "proposed remedial framework for big data privacy", designed to guide future research in the evolving domain of big data privacy.

Research Method and SLR

Considering the criticalness of the topic, this study adopted a systematic literature review to investigate the specific challenges associated with Big Data privacy. Published articles were collected from distinct academic journals, publisher platforms and professional sources, including IEEE Xplore, Springer, MDPI, Taylor and Francis, ACM Digital Library, SSRN, ArXiv, Computers and Security, IAPP and other indexed journals. Further literature was obtained through institutional access, Google Scholar, Google AI, Google Search, Mendeley and fruitful guidance from the research supervisor to ensure broad coverage of relevant studies.

A structured search strategy was applied, in addition to multiple related terms and keywords with Boolean operators to refine overall results. Search terms include big data privacy, data security, big data privacy challenges, regulations (PDPL, GDPR, CCPA), Compliance and Personal Data Protection. Both journal articles and conference papers published between 2000 and 2025 and only articles written in English were considered. Studies that did not directly focus on big data, data privacy, data security and data compliance or lacked sufficient academic quality were excluded. Repeated issues were systematically recognized and grouped into major findings. This structure was polished through iterative refinement to guarantee precise and consistent framework.

Limitations include the synthesis of information from multiple scholarly articles, official websites and the author's independent analysis regarding regulatory frameworks. Its purpose is to assist future research and contribute to resolving industry related privacy issues. However, researchers and reviewers are advised to

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

independently revalidate and verify all findings and interpretations presented herein. Wherever applicable, efforts were made to reference primary sources at the time of citation to enhance the authenticity, validity and credibility of this study.

Discussion and Reflection

Big data

Generally Big Data refers to data that is very large, fast and complex, commonly described by the 3Vs i.e. Volume, Velocity and Variety. According to IBM, there are 4Vs i.e. Volume (scale of data), Velocity (speed of data in/out), Variety (forms of data) and Veracity (reliability of data). Oracle expands the definition to 5Vs i.e. Volume, Velocity, Variety, Veracity and Value.

In today's interconnected digital era, Big Data has become one of the most invaluable assets in the world, a core driver of emerging technologies, business growth and innovation. Therefore, in today's Data Driven society, almost every one activity leaves behind a digital transformation. Whether someone is browsing the internet, using GPS for navigation purposes, watching videos, being involved on social media, or simply making an online purchase, data is being constantly created, collected, analyzed and ultimately monetized.

As illustrated above the Big Data ecosystem includes multi-dimensional interconnected layers, including (but not limited to) multi-modal data sources, ingestion pipelines, distributed data storage, processing frameworks, analytics, applications and visualizations (Figure 1). While this layered structure enables organizations to derive insights from vast and varied data, it also introduces significant complexity and unavoidable challenges. The constant influx of diversified data sources, ranging from IoT sensors and mobile applications to social media streams and transactional systems, creates significant challenges in ensuring seamless integration, scalability and reliability. Moreover, the coexistence of batch and real-time processing demands advanced orchestration techniques to maintain data consistency and timeliness. Beyond technical integration, the ecosystem's complexity extends to issues of governance, security, as well as privacy, as organizations must enforce compliance and safeguard sensitive information across heterogeneous environments. The big data most of the time contains PII (Personally Identifiable Information) like names, addresses, contact details, financial records, health data and sometimes behavioral patterns and therefore, the concept of data security and privacy has become crucial and storage technologies are continuously improved to provide better security [1]. All together, these factors illuminate why managing the Big Data ecosystem remains a vital and ongoing challenge for both researchers and practitioners.

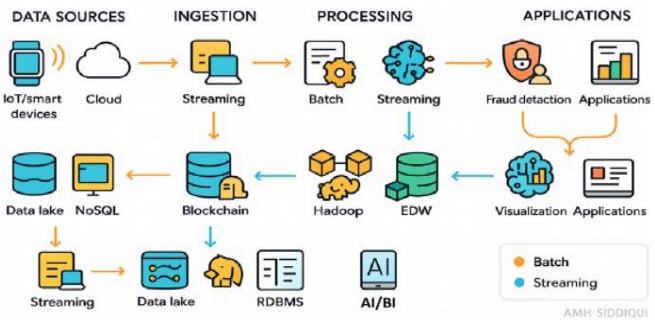

Figure 1: Big data ecosystem architecture.

To understand global data utilization, social media platforms such as Facebook, YouTube, Instagram and LinkedIn collectively engage billions of users worldwide, with Facebook alone reporting over 3.07 billion active users, while every single day, YouTube gets about 720,000 hours of new content (Figure 2). The massive scale of user interactions, likes, comments, shares, video uploads, searches and professional networking activities, demonstrates how quickly distinct form of data is generated on a global scale. Every second, these platforms produce vast amounts of structured and unstructured data, contributing significantly to the velocity (rapid growth), volume (massive amount) and variety (complexity) that highlight the complexity of Big Data. Managing this non-stop inflow of user-generated content certainly entails infinite challenges for organizations, as it requires auto-scalable storage solutions, ultra-smart processing frameworks and strong AI based governance mechanisms to guarantee data quality, security and necessary privacy compliance.

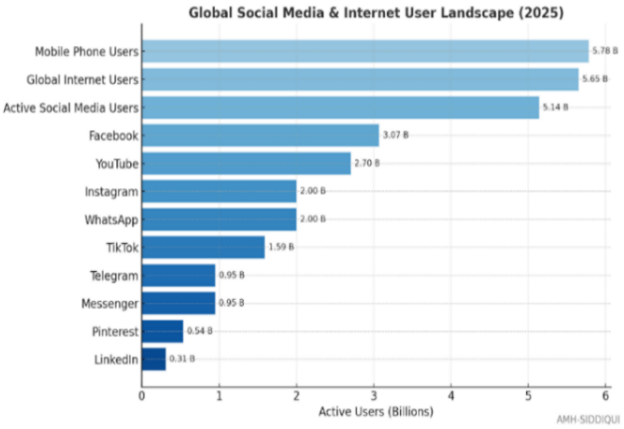

Figure 2: Global digital usage statistics.

Data privacy

In a nutshell, Data privacy is nothing more than right and practice of keeping personal data safe, controlling authorized access and managing legitimate data utilization. Data privacy has become a cornerstone of the modern digital landscape, often referred to as the "new currency" in the digital age [6]. Any breach - The breach included unauthorized access to credit card information, numbers, expiration dates and CVV codes and in certain cases, information about the identity of the cardholders, all obtained through hacking and malware of sensitive data may lead to severe negative consequences, including identity theft, financial fraud, or even physical harm [7]. For example, stolen banking credentials can lead to unauthorized withdrawals, while leaking health information might affect insurance coverage or employment opportunities.

Maintaining control over personal data empowers individuals to decide who can access their information, how it is used and under what circumstances. In addition to individuals' self-respect, freedom, peace of mind and enhancing trust in digital services, this control also prevents annoying marketing, scams and unauthorized surveillance. Moreover, GDPR (General Data Protection Regulation) mandates that individuals have the right to understand how their data is processed [8]. Under GDPR, consent, transparency, purpose limitation and data minimization are four essential principles, ensuring lawful and fair data processing.

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

- User Consent: Getting clear permission from users before using their data, while balancing data utilization, data security and data privacy. Balancing privacy and security are complex, involving factors like user consent and data retention. Privacy-enhancing technologies aim to strike this balance [4].
- Data Minimization: The collected data need to be adequate, relevant and limited to what is necessary in relation to the purpose of processing [9]. Collecting only the data that is really needed.
- Purpose Limitation: Personal data needs to be collected for a specified, explicit and legitimate purpose [9]. Data should only be collected for specific, explicit and legitimate purposes and not used for anything beyond those purposes without further consent.
- Transparency: Transparency is equally vital, particularly in industries like healthcare and finance, were data management decisions impact lives and livelihoods [8]. Clearly informing users about data use.

Big Data Privacy

It refers to the issues, challenges and techniques around protecting individuals' privacy when dealing with large-scale datasets, often aggregated from many diversified and multi-model sources, which mostly include sensitive personal information. Because of the volume, velocity, variety and interconnection of such data, traditional privacy measures become insufficient and require significant adaptation. The growth of AI - Artificial Intelligence and ML - Machine Learning, has created new privacy challenges. These technologies can uncover and reveal hidden characteristics or behaviors from everyday data, even allowing anonymized information to be traced back to individuals. Indeed, big data has advanced faster than traditional privacy protections and older methods like encryption, data anonymization, access control and data minimization are no longer enough on their own. At the same time, "The advent of emerging technologies, such as quantum computing, edge computing and biotechnologies, poses unprecedented challenges to existing privacy frameworks" [10]. To highlight this transformation, Table 1S compares key characteristics and outcomes of data practices before and after Big Data adoption. As a result, the demand for privacy-by-design, privacy-by-default architecture, ethical data governance frameworks and real-time data protection mechanisms has never been higher. According to Mulder and Vellinga, every stakeholder should apply the principles of privacy by design and privacy by default [9].

Big data privacy dimensions

There are four primary dimensions, where maintaining data privacy is both essential as well as challenging. Each domain entails exclusive constraints, requirements and complexities. Deploying effective and smart big data privacy mechanisms across all these dimensions is imperative to prevent negative consequences such as data breaches, legal penalties, loss of public trust, loss of credibility and ethical violations.

- **Legal:** Different regions obligate distinct levels of data protection regulations. For instance, GDPR imposes strict clauses relating to data processing, user consent and cross border transfers. In comparison, the CCPA (California Consumer Privacy Act) focuses more on transparency and consumer rights, while Saudi Arabia's PDPL (Personal Data Protection Law) emphasizes data localization and prior consent for international transfers. As a result, multinational companies must therefore navigate incompatible legal standards, creating uncertainty and compliance challenges across jurisdictions.
- **Technical:** Maintaining data privacy within complex digital infrastructures like Big Data ecosystem, persists remarkable challenges. Despite legal compliance, technical missteps, for example, misconfigured cloud servers or weak encryption protocols, can result in massive data breaches.
- **Political:** Governmental regulations and data control policies often create challenges for global privacy operations because countries apply different rules. For example, international companies such as MDM-Team and Bilim-Base operate in many countries. Sometimes they may require transfer of user data to their global servers from one country to another. However, the country's new law restricts this practice, stating that "Our country users' data cannot be freely transferred abroad."
- **Ethical:** Beyond legal, political and technical compliance, respecting user autonomy, ensuring informed consent and minimizing data collection reflect moral accountability rather than only regulatory obligation. Ethical data management builds public trust and supports the principle that privacy is a fundamental human right, not just a compliance requirement.

Big data privacy on four major layers

As illustrated in Figure 3, privacy can be protected through a multi-layered approach (i.e. Management, Interface, Storage and Access layers) together to secure information from unauthorized access and misuse. Each layer plays a unique role in ensuring that personal and organizational data remain safe, secure and confidential.

- **Management layer:** This layer covers the policies, processes and people that govern and control data, privacy reviews, data inventories, vendor oversight and incident response. Embed privacy-by-design, track lawful bases and consent. For example, before launching a new enterprise analytics feature, DPIA flags birthdate collection as unnecessary. Therefore, in accordance with Data Minimization principle of GDPR, the team removes it and updates the record of processing.
- **Interface layer:** This layer primarily used distinct APIs and forms for data acquisition. Aiming to avoid dark, misleading patterns and excessive collection, use clear, granular consent, minimize forms' fields and provide contextual notices. For instance, in the sign-up page ask only for email and password, mark extra profile fields as optional, or turned off by default and link a concise, action specific privacy notice.
- **Storage layer:** This layer is known as data repository, or data reservoir for databases, file shares, backups and logs. Against theft, leaks and over retention, apply state-of-the-art encryption at rest with strong key management, minimize what you store and enforce short period of retention. As an example, a company should store customer PII in an encrypted database with keys in KMS and automatically purges access logs after 30 days.
- **Access layer:** This layer manages and controls who and what can use data at runtime. Preventing improper use *via* strong authentication mechanism and by applying principle

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

of least privilege authorization. Enforce MFA, passkeys, short lived tokens, just-in-time access and detailed immutable audit logs. For instance, authorized support staff, having legitimate access, must request time-boxed (e.g. 15-minute) access to a customer record with manager approval; every activity must be properly logged.

Figure 3: Big data privacy – major layers.

Major components of big data privacy attack

Figure 4 frames data privacy threat assessment centered on four interconnected components i.e., Adversary, Motivation, Target and Methodology. They all together describe who (malicious attacker is attacking), why (intention and reasoning behind the seen), what (aim to impact) and how they do it (approach and technique). For instance, a small hacktivist group (Adversary) with only moderate IT skill sets aimed at selling PII (Motivation) without a lawful basis. They planned to target the passengers' profile database and the organization's Active Directory (Target) to obtain PII (Personally Identifiable Information) for sale. They used phishing emails with catchy subject lines, to trick users into giving up credentials for privileged accounts, then moved silently through the network while covering their tracks to avoid detection (Methodology).

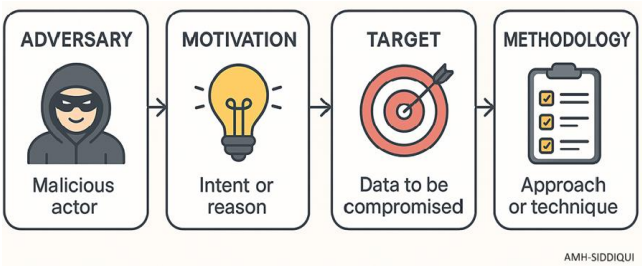

Figure 4: Big data privacy attack - major components.

Subsequent to privacy assessment and identifying the adversary, target, methodology and motivation, the next step is to prioritize and validate the findings by ranking them according to severity. Immediate mitigation must be applied to any active threats, such as isolating affected systems, revoking access, or patching vulnerabilities. Then, a concise report should be prepared, summarizing the scope, key findings, recommended fixes and responsible teams. Once fixes are implemented, lessons learned should be documented and security policies, monitoring and response plans should be updated to strengthen future defenses. Rule of thumb: Adopt Privacy-By-Design principle and uphold a resilient, proactive security posture.

Having discussed the fundamentals of four major Big Data dimensions where privacy is indeed mandatory, the four major layers where it can be effectively protected and the four key components where privacy threat assessment is primarily focused, it is crucial to turn attention toward the key threats associated with data privacy in the Big Data landscape. Subsequently, the following section explores the core obstacles, highlighting how regulatory gaps, technological limitations and behavioral factors continue to interfere with the realization of genuine data confidentiality and user trust.

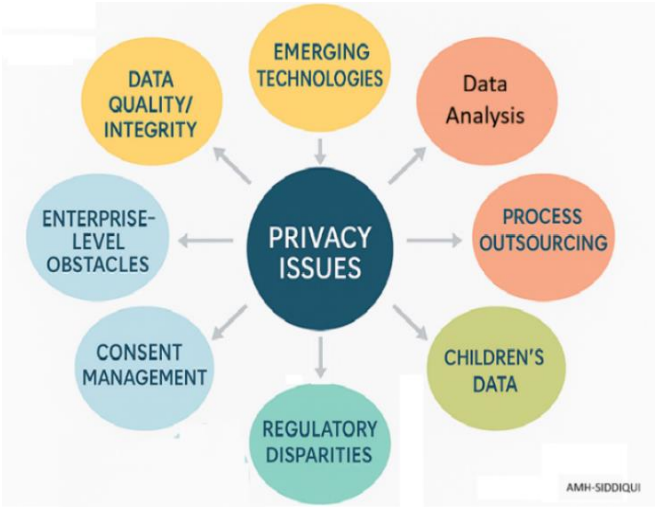

Figure 5: Big data – great privacy challenges.

Identified emerging challenges in the light of big data privacy

The study identified twelve core challenges and approximately forty-four sub-challenges related to data privacy within the big data ecosystem. For instance, among the core challenges (illustrated in Figure 5), such as maintaining data quality and integrity, emerging technologies and enterprise-level privacy governance, regulatory disparities across borders and data analytics, each encapsulating multiple layers of issues and complexities. Technological factors such as the rise of IoT, Blockchain, AI and FL (Federated Learning) introduce additional risks of data exposure, while organizational barriers such as fragmented privacy structures and limited resources, further destabilize effective governance. Collectively, these twelve core and forty-four sub-challenges illustrate that Big Data Privacy is not limited to cyber threats alone but extends across ethical, operational and policy related dimensions, demanding a holistic and adaptive privacy management framework.

Maintaining Data Quality and Integrity

Data quality

Data quality refers to ensuring that personal data is accurate, complete, current and relevant for the purpose it was collected. Poor

data quality creates privacy issues not because the data was tampered, but because it is incorrect or outdated. For example, if an organization stores outdated addresses, sensitive letters might be sent to the wrong person, exposing private information. Unlike data integrity, which focuses on preventing unauthorized changes or corruption, data quality is a governance issue, requiring regular audits, updating policies and validation checks to ensure data remains trustworthy and fair in decision-making processes.

Data integrity

Represents another major privacy challenge in the context of Big Data due to the massive volume, variety and velocity of information collected from numerous interconnected, multi-model sources. As data moves across multiple platforms and stakeholders, ensuring that it remains accurate, consistent and unaltered throughout its lifecycle becomes increasingly complex. Any distortion or manipulation of data not only undermines trust and weakens analytical reliability but can also expose individuals' private information through mismatched or corrupted records. Moreover, the integration of heterogeneous data streams, such as those from healthcare, financial systems and social networks, amplifies the risk of unauthorized alterations or linkage attacks that threaten both integrity and privacy. Therefore, deploying robust integrity framework, including encryption, immutable audit trails and data validation mechanisms, is vital to preserve privacy and maintain confidence in insights derived from big data. This challenge becomes even clearer when examined through below-mentioned examples, considering different sectors.

Data Integrity Breach Factors

Cybersecurity threats

Cybersecurity threats occur when hackers, malware, or viruses gain unauthorized access to Big Data Ecosystems. Once inside, they can manipulate, steal, or erase critical data, making it unreliable and unsafe for use. This directly threatens data privacy, as sensitive or personal information may be altered, or exposed to outsiders without consent. If hackers penetrate a hospital's database and alter a patient's blood type from "O+" to "B+," it could endanger the patient's life, cause breaches of confidentiality and violate data protection regulations, including HIPAA and GDPR. Ensuring data integrity requires strong security mechanisms, including MFA (Multifactor-Authentication), up-to-date antivirus programs, immutable auditing, data masking and state-of-the-art cryptographic methods such as PET (Privacy-Enhancing Technologies).

Data transmission failures

When data travels through the internet or intranet, weak connections or lack of in-transit encryption can result in data loss, alteration, or interception. This becomes a data privacy challenge because private or confidential information might be exposed or corrupted during transfer. Imagine a company sending employee salary information to another branch. If the network is not secure and part of the data is lost or changed, salary details may be misrepresented or leaked and compromising employee privacy. Always use secure, encrypted communication channels such as HTTPS, SSL or VPNs and verify that transmitted data matches the original before use.

Human error

Human blunders such as incorrect data input, deleting files accidentally, or sending data to the wrong recipient, can severely harm both data accuracy and privacy. Since human actions are unpredictable, privacy breaches often occur unintentionally, making them one of the most common data privacy challenges. An employee mistakenly emailing a list of customer emails to the wrong person exposes private contact details to unintended parties. Regular employee training, rechecking mechanism and following PoLP (Principle of Least Privilege) can lower the likelihood of errors.

Lack of standardization

Lack of standardization means inconsistent data formats or privacy practices, which can cause misinterpretation, erroneous delivery, or exposure of personal information. For instance, if one office records "01/04/2025" as April 1 and another as January 4, merged data may contain reporting errors or reveal wrong details to the wrong recipient. Inconsistent email, naming conventions can also lead to privacy breaches. For example, a hospital's HR uses misiddiqui@ourhospital.com for Muhammad Imran Siddiqui, while the lab uses msiddiqui@ourhospital.com. Lab results are mistakenly sent to Maryam Siddiqui, exposing confidential patient data, a violation of HIPAA or GDPR. Similarly, in a bank, payroll emails salary slips to msiddiqui@ourbank.com instead of misiddiqui@ourbank.com, leaking financial details to another employee. Likewise, when two companies merge and follow different email formats, HR may send performance or medical records to the wrong individual(s), exposing PII such as salary and health information. Even though the root cause is inconsistent data standards, the result is a data privacy breach. Adopt standardized data formats, consistent privacy guidelines and universal protocols across all departments to maintain clarity and prevent unintended data exposure.

Sensor malfunctions

In Big Data Ecosystem, sensors and IoT devices collect real-time data. If these sensors malfunction or provide incomplete data, they can lead to false interpretations or unintended sharing of private information. As a result, it may lead, data privacy challenge because users may lose control over what personal data is being recorded or transmitted inaccurately. A faulty heart rate sensor in a fitness tracker may send incorrect health information to a mobile app, which can share it with other platforms, violating user privacy. Conduct regular maintenance and calibration of sensors, apply data validation checks and restrict automatic data sharing until the information's accuracy is confirmed.

Data quality and integrity - contextual relevance to the aviation big data ecosystem

In the aviation industry, vast amounts of data are continuously generated from multi-model sources such as flight operations, maintenance logs, passenger bookings, radar systems and IoT sensors embedded in aircraft. Maintaining data quality and integrity is therefore crucial, as inaccurate or corrupted data can lead to serious consequences. For example, inconsistent flight path data or incomplete engine sensor readings may misrepresent predictive maintenance analytics, potentially cause delayed repairs or overlook faults. Similarly, inaccurate passenger identity data can compromise border security or violate privacy laws like PDPL and the EU's

GDPR. The challenge becomes more complex when integrating data from multiple airlines, airports and third-party service providers, where differences in data formats, validation standards and collection ethics create additional risks to data accuracy and trustworthiness.

Occurrence probability (High - 0.6 to 0.8)

Given the large scale and velocity of aviation data, the probability of data quality and integrity issues occurrence is high. The aviation ecosystem depends heavily on real-time data exchange between multiple systems like air traffic control, airline operation centers, weather forecast and airport logistics, each with its own data governance practices. Issues may occur due to sensor malfunction, delayed synchronization, or manual entry errors. For example, if a single sensor on a Boeing 787 transmits erroneous altitude data due to software lag, it could propagate through the analytics pipeline, misleading flight planners or performance evaluators. Furthermore, the continuous use of legacy systems in aviation increases the likelihood of data duplication, loss, or misclassification when integrated with modern cloud-based analytics platforms.

Consequential impact - data quality deficiencies and compromised data integrity

Poor data quality and compromised integrity in aviation can have severe operational, financial and reputational impacts. Inaccurate maintenance data may cause flight delays, unexpected breakdowns, or even safety hazards if issues are not detected and resolved in time. On the privacy front, altered, corrupted or unverified passenger data can result in identity mismatches, unauthorized access, or compliance breaches under stringent privacy laws. Loss of trust in the accuracy of operational data can also damage airline credibility and regulatory compliance, especially in jurisdictions with strict oversight.

Emerging Technologies - The Dark Side of Innovation

The growing significance of the Big Data Ecosystem and the importance of data privacy, it has proven that technological progress has brought about remarkable opportunities for innovation, efficiency and insight. However, it has also introduced several unavoidable challenges related to data protection, ethical use and individual rights. As per Oluwabunmi Layode et al., [11]. “The advancement of technologies such as remote accessibility, data interchange mining and cloud computing has introduced unprecedented challenges to privacy and protection”. To better understand these challenges, it is essential to examine how certain key technologies pose distinct privacy and security threats within the broader data ecosystem. Each of these technologies introduces unique vulnerabilities that demand careful analysis in addition to robust data privacy framework. According to Debbarma “The evolution of technology has outpaced the development of privacy laws, necessitating a fundamental reassessment of legal frameworks to address emerging challenges effectively”.

Telecommunication

It is one of the primary privacy challenges, because it involves transmitting personal and sensitive information over networks that can be intercepted or hacked easily, if required security measures are not in place. Protecting calls, messages and data from intrusion and unauthorized access is difficult but indispensable. If a hacker breaks into a mobile network, they could listen to private phone calls or steal text messages, putting users’ privacy at risk.

IoT – Internet of Things

IoT refers to the interconnected network of physical devices (sensors, cameras, appliances, vehicles, etc.) that communicate and exchange data over the internet. According to Ijaiya [6]. “Emerging technologies, including the Internet of Things (IoT) and biometric systems, present new privacy challenges that demand urgent attention. IoT devices, such as smart home assistants, wearables and connected cars, continuously collect data to enhance user experiences. However, the pervasive nature of these devices has led to concerns about ubiquitous surveillance. For instance, smart speakers like Amazon Echo and Google Home have faced criticism for recording and storing conversations without user consent”.

Biometrics and surveillance technologies

Biometrics Systems (such as fingerprint scanners, facial recognition, voice authentication, steps or walk recognition etc.,), that identify or authenticate individuals using unique biological or behavioral attributes. Biometric data, used in facial recognition systems, fingerprint scanners and genetic testing, introduces additional complexities. According to Ijaiya [6]. “In law enforcement, facial recognition technology has been deployed to identify suspects, but it has also raised concerns about racial biases and wrongful arrests”. On the other hand, Surveillance tech (such as CCTV, drones, license plate recognition, AI-based video analytics etc.) includes tools and systems explicitly designed to observe, monitor, track, or record people, environments, or activities. As per Oluwatosin Reis et al., [10]. “Biometric data, being unique and immutable, raises concerns about the potential for identity theft and the irreversible consequences of a breach. Moreover, the mass collection and storage of biometric data without robust safeguards can lead to unwarranted surveillance, compromising individuals' privacy rights. Addressing the privacy challenges posed by biometrics and surveillance technologies requires careful legal considerations and the establishment of effective”.

Adoption of cloud computing

As stated by Goel et al., that “Only the large-scale infrastructure of the cloud is capable to manage this bigdata [12]. The security and privacy of cloud infrastructure are a challenge for the diversified storage of data”. Using multiple cloud providers or a hybrid infrastructure (mix of cloud and on-premises) creates privacy challenges because data moves between different systems with different security rules. This makes it harder to control who can see or use the data, increasing the risk of leaks or unauthorized access. “As companies adopt cloud services, data protection becomes more complex: Companies may not know where all applications and data are stored; Third Party hosting limits visibility into data access and sharing; and Shared security responsibilities may be misunderstood or misapplied” [3].

Privacy-preserving techniques

As reported by Shahriar et al., “each privacy-preserving technique comes with its own set of challenges and limitations [13]. These limitations often revolve around computational costs, scalability, trade-offs between privacy and utility and the ability to address specific privacy risks effectively”. Most of the organizations assume

https://doi.org/10.70844/ijas.2025.2.42

that once they apply privacy-preserving techniques, the data is fully safe. Even though PPT such as differential privacy, anonymization, federated learning, encryption and homomorphic computation, are designed to protect data, they can produce new privacy vulnerabilities when misapplied, misconfigured, or improperly implemented. For instance, weak aggregation settings in federated learning can still leak sensitive information.

Adoption of sophisticated encryption

According to Oluwatosin Reis et al., "the widespread adoption of sophisticated encryption techniques may also pose challenges for law enforcement agencies seeking access to encrypted information for legitimate purposes" [10]. Additionally, different encryption techniques more often bring significant computational overhead, impacting overall performance in real-time applications. Moreover, managing balance between data utilization and encryption strength, is a constant major concern, specifically for small-scale organizations with limited resources.

Differential Privacy (DP)

DP Introducing controlled noise to data to safeguard individual identities. Goel, stated that "A differential privacy approach is presented to preserve privacy in big data [12]. The model is designed in such a way that ensures an equal possibility of data to get released amongst all input data. Two mechanisms known as the Laplace mechanism and the Exponential mechanism are designed for providing differential". While Williamson and Prybutok, say: "Differential Privacy (DP) has become an indispensable tool in the realm of healthcare data analysis, offering a robust solution to the perennial challenge of maintaining patient confidentiality while still allowing for the extraction of valuable insights from large datasets [14]. Additionally, several organizations including Tech-Giants like Apple and Google, have adopted differential privacy to protect user data in real world applications, but according to Hakeemat Ijaiya, "differential privacy faces challenges in maintaining accuracy for complex queries and balancing noise levels with analytical utility". Moreover, "Excessive noise can obscure critical insights, while insufficient noise may compromise privacy" [8].

Federated Learning (FL)

As mentioned by Hakeemat Ijaiya, "Federated learning is an innovative AI approach that enables machine learning training across decentralized datasets without transferring raw data", aiming to protect user privacy. Whereas, this reduces the risks of exposing sensitive data, FL still faces privacy challenges because model updates can indirectly leak information and malicious participants can poison or manipulate the system and even attackers may infer user identities or behaviors. As a whole, "The benefits of federated learning extend beyond privacy preservation to include reduced data transfer costs and improved data security" [8].

Wearable devices

As mentioned by Zhou Sogand Hasanzadeh and Zhou, "While wearable devices could benefit construction workers in many aspects, potential challenges still exist [15]. Assume your employer is asking you to wear a device at work. To what extent do you have privacy concerns if employer provided wearable devices collect your mental state data (e.g., stress)?". Several wearable devices are connected to mobile apps and cloud services, which means the data can be shared to third parties, including marketing agencies and/or insurance companies for profiling and advertisement purposes. If this data is shared without individuals' explicit consent, it cannot only expose personal details like health and lifestyles, but it is also violation of privacy regulations like PDPL and GDPR.

Hadoop (HDFS and map reduce)

Hadoop Processes large scale distributed datasets in an efficient and secure manner. Hadoop has two core components, namely Map Reduce and HDFS (Hadoop Distributed File System). "A Map Reduce first divides the data into individual chunks which in turn are processed by Map jobs in parallel" [16]. HDFS stores large datasets across multiple computers also known as nodes. In Hadoop HDFS, "A mechanism of trust is implemented between the name node and the user that is a component of HDFS." [12]. This setup can become a privacy challenge if not secured properly. For example, if a user's identity is not strongly verified, someone could pretend to be another user and access sensitive files like financial records stored in the system, leading to unauthorized data exposure. Eventually, proper privacy and security of distributed framework like Hadoop is essential and is a challenge because the function like MapReduce is used for mapping of the data and HDFS creates multiple replicas of each data block for fault tolerance. Although Hadoop technology increase's reliability, in parallel, it increases risks, expands the attack surface and complicating data deletion (e.g. under PDPL "Right to be Forgotten").

Blockchain technology

Oluwatosin Reis (2024) [10] says, "Blockchain technology, known for its decentralized and secure nature, has implications for identity management and data ownership. Decentralized identity systems empower individuals with greater control over their personal information, potentially reducing reliance on centralized authorities". Williamson and Prybutok, mentioned, [14]. "This dynamic blockchain ecosystem is not just a static data repository but a transformative force that brings unprecedented transparency, efficiency and reliability to healthcare data management". Indeed, Blockchain Technology is regarded as a privacy-preserving technique; however, it entails the following challenges:

- Data Exposure in Blockchain: In Blockchain, data is stored in a series of connected BLOCKS. Each block contains a list of transactions and once a block is full, it is linked to the previous one, forming a CHAIN. The integrity of this chain is maintained by the network of connected computers, called NODES. If Blockchain is not encrypted well using state-of-the-art mechanism, sensitive data may be exposed to all nodes.
- Immutability Risk: Blockchain is a digital record keeping system. Once private information is added, even mistakenly, it can't be removed, as records are uploaded, stay forever, to any extent, it cannot be changed.
- GDPR and PDPL Compliance Issues: According to Williamson and Prybutok [14]. Right to Erasure (Right to be Forgotten) is not possible on blockchain (data is permanent). While most of the data privacy regulations impose stringent requirements on personal data processing, focusing on data subject rights such as consent, the right to access, rectification and the right to be forgotten.

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

- Scalability: Blockchain technology is designed primarily to store big data. Blockchains are slow and not ideal for big volumes of personal data.

Online tracker technology

Trackers follow individuals across websites and applications to enable profiling based on interests and habits. For instance, during your visit to baby name website, site stores tracking files and sensitive insights, sometimes without consent. The Ad-System (like Google or Meta Ads) assumes you might be a soon, to be parent and automatically shows pregnancy or baby related Ads. Tracking also extends to schools and workplaces, watching what people act and say, limiting free expression. Hidden tracking often violates privacy laws like GDPR or CCPA and shadow profiles persist even for users who never sign up, without their consent. Email beacons track when and where messages are opened, while session replay scripts record clicks and keystrokes, risking leaks of private data, causing data privacy violation. Anonymous data can often be re-identified by combining location, device model and timestamps (e.g., map pings that match your home and work pattern), causing a breach of privacy compliance, especially when person wants to remain anonymous.

Increasing reliance on AI, XAI and ML

AI (Artificial Intelligence) is the broad field focused on creating machines that can mimic human intelligence like self-driving cars that navigate traffic or virtual assistants such as Siri and Alexa understand your voice. Within AI, ML (Machine Learning) allows systems to learn from data and improve over time, much like banks detecting fraud from transaction patterns. To ensure trust and transparency, XAI (Explainable AI) steps in by revealing why an AI makes certain decisions such as showing that a loan was declined due to low credit score or insufficient income, making advanced AI systems not only powerful but also accountable and understandable. Insofar AI and ML utilization is concerned, Taherdoost, mentioned [17]. “Big data and machine learning are being used more and more frequently across a wide range of industries, including healthcare, transportation, athletics and finance. Although these technologies have many advantages, they also raise ethical and privacy issues that should be considered when making decisions”.

Figure 6 is designed to present these differences in an intuitive image format, enabling viewers to quickly grasp how each concept fits within the broader AI ecosystem. According to Oluwatosin Reis et al. “The increasing reliance on artificial intelligence and machine learning in decision-making processes introduces challenges related to algorithmic transparency, bias and accountability”. Moreover, “The advent of Artificial Intelligence (AI) and Machine Learning (ML) has ushered in a new era of data processing capabilities, enabling systems to analyze vast amounts of information with unprecedented speed and accuracy. While these technologies offer tremendous benefits in areas such as healthcare, finance and transportation, they also raise profound concerns about individual privacy.”. It was also emphasized by Hakeemat Ijaiya, that “AI Driven privacy tools are increasingly targeted by adversarial attacks, which exploit vulnerabilities in AI models to compromise data confidentiality” [8]. Furthermore, “The risk of unauthorized access, misuse and the potential for AI Driven decisions impacting individuals' lives without transparency pose significant challenges to privacy in the digital age” [10]. During machine learning, many privacy risks exist, like data leaks before or after training; models unintentionally revealing personal data and attacks that guess private information from the model. So, the work to protect privacy in ML is still young and developing. More research and better methods are needed to keep data safe before, during and after the ML process. Dwivedi et al., highlighted, “In response to the privacy implications of AI and ML, regulatory bodies worldwide are actively engaged in developing frameworks to address these challenges”.

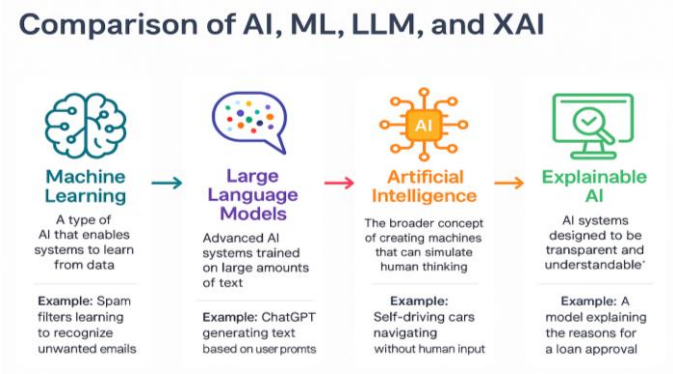

Figure 6: AI, ML, LLM and XAI – A momentary look.

LLM - Large Language Model

LLMs by learning the structure of human language; they can hold natural conversations, summarize complex texts, or even write creative essays. ChatGPT and Google Gemini are the perfect examples for LLM. As LLM are powerful artificial intelligence systems trained on massive volumes of text data, these datasets often include personal, sensitive, or proprietary information, present serious privacy challenges. During training, models may inadvertently store fragments of such data and later reproduce or expose them. In essence, the process that allows LLMs to generalize from large datasets can also make them prone to leaking private information. For individuals or organizations deploying these systems, it creates a persistent challenge between maximizing usefulness and ensuring the confidentiality of the underlying data. Below are six major categories of privacy risks relating to LLM:

Emerging technologies - contextual relevance to the aviation big data ecosystem

The aviation industry is undergoing a profound digital transformation driven by the adoption of emerging technologies such as AI (Artificial Intelligence), IoT (Internet of Things), Blockchain, Biometrics and Surveillance. These innovations are being integrated across operational, maintenance and passengers’ service domains to enhance safety, efficiency and overall travel experiences. Modern aircraft and ground systems now generate and process enormous volumes of real time data from interconnected sources, ranging from aircraft sensors and air-traffic management systems to passenger mobile applications and biometric checkpoints. For instance, next-generation jet engines equipped with IoT sensors can transmit several terabytes of operational data per flight to centralized maintenance hubs for predictive analytics and fault detection. Similarly, biometric identification systems, such as facial or fingerprint recognition at airports, are streamlining passenger processing and improving security but simultaneously capturing highly sensitive personal data. Blockchain platforms are increasingly being explored for secure data sharing, ticket validation and supply-chain transparency, while AI and machine-learning models are being leveraged for route optimization, weather prediction and fraud detection.

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

Occurrence probability (Very High - 0.8 to 1.0)

The probability of privacy breaches arising from emerging technologies in aviation is very high and always increasing. The growing interconnectivity of digital systems such as aircraft-to-ground communication networks, automated baggage handling and in-flight Wi-Fi creates numerous data exposure points. For instance, a single misconfigured cloud storage bucket containing passenger manifests or maintenance logs could lead to a large-scale data leak. The use of third-party AI and analytics solutions also raises risks of data misuse or poor anonymization practices. As the aviation sector accelerates its digital transformation, with over 90% of airlines adopting cloud-based analytics and biometric systems, the likelihood of privacy incidents is expected to rise unless unbreakable privacy frameworks are embedded early in the technology lifecycle.

Consequential impact - deviation from proper technological practices

A data privacy breach in aviation can have severe operational, financial and reputational consequences. Exposure of passenger data can lead to violations of data protection laws such as PDPL and GDPR, result in heavy regulatory fines. Voss, highlighted that "Vueling Airlines received the first actual administrative fine under the GDPR covered by this study in the amount of €30,000, on October 1, 2019, issued by the Spanish supervisory authority Agencia Española de Protección de Datos (AEPD)" [7]. According to Shahriar et al., "Some notable examples of GDPR fines include the €50 million fine against Google and £183 million against British Airways" [13].

Collaborative Data Processing

Collaborative data processing, where multiple organizations or systems jointly handle and analyze data, creates major privacy challenges because it extends data usage beyond the control of any single entity. In healthcare, for instance, hospitals may collaborate to train AI models for early cancer diagnosis, keeping patient data locally but sharing model parameters that can still reveal private medical details. In the financial industry, banks working together on fraud detection might unintentionally reidentify clients by combining anonymized datasets. Likewise, in manufacturing, companies coordinating on supply chain analytics can expose confidential pricing or vendor information through shared insights. In education, universities pooling student performance data for adaptive learning tools may risk breaching student confidentiality if anonymization is insufficient. These situations highlight how, when data moves across organizational boundaries, threats like data leakage, inference attacks, loss of control and regulatory breaches become more likely, demanding strong encryption, strict access controls and well-defined legal and ethical frameworks to ensure data privacy in collaborative environments.

Collaborative data processing - contextual relevance to the aviation big data ecosystem

In the aviation industry, collaborative data processing has become central to modern operations, from sharing passenger data across airlines and airports to integrating aircraft maintenance logs, air-traffic information and weather data among global aviation networks. Airlines increasingly collaborate with partners such as air-traffic control authorities, airport operators, customs agencies and third-party analytics firms to enhance safety, efficiency and passenger experience. For instance, collaborative platforms like the International Air Transport Association's (IATA) "Aviation Data Exchange" promote shared insights on flight routes, emissions and maintenance schedules. However, these collaborations also create significant privacy concerns, as large volumes of sensitive data including passenger identities, biometric data and operational security logs, are pooled and processed across multiple jurisdictions. This increases the risk of unauthorized access, misuse, or data breaches, especially when privacy frameworks differ between partners or countries.

Occurrence probability (High - 0.6 to 0.8)

The probability of privacy challenges evolving from collaborative data processing in aviation is high and growing. This is due to the increasing reliance on digital transformation initiatives such as predictive maintenance using IoT sensors, real-time passenger analytics and AI-based air-traffic coordination systems. Each new layer of technological integration adds more data-sharing nodes and thus more potential points of vulnerabilities. For example, the use of cloud-based systems connecting airlines, booking platforms and government immigration databases heightens the risk of inadvertent data exposure if access controls are weak or inconsistent. Moreover, the aviation sector operates on a global scale, involving partners with varying levels of cybersecurity maturity and compliance with data protection regulations such as KSA PDPL, GDPR Privacy Framework. This heterogeneity makes privacy breaches during joint data processing a likely and recurring threat.

Consequential impact - non-compliant collaborative frameworks

The negative consequences of compromised privacy in collaborative data processing within aviation can be severe and multi-dimensional. On an operational level, a data breach could lead to grounded flights, disrupted schedules and loss of trust between international partners. On a regulatory level, Non-Compliance with cross-border privacy laws (such as GDPR) could trigger heavy penalties and fines. The reputational damage may be even more costly than seen in incidents where airlines suffered passenger criticism following cyberattacks or accidental data leaks. Beyond financial and reputational losses, privacy failures in collaborative processing could jeopardize national security by exposing sensitive data such as flight paths, cargo manifests, or diplomatic travel details. Thus, the impact extends beyond individual organizations to the entire aviation ecosystem.

BPO - Business Process Outsourcing

According to Sajid Momin et al., "BPO is Business process outsourcing in IT/ITES industries" [3]. BPO is one of the greatest privacy challenges, because sensitive customer and business information are handled on daily basis, by third party companies. This heightens the risk of unauthorized access, data theft, or misuse of personal details. Weak privacy controls, untrained staff, or poor security measures can lead to serious and unavoidable breaches. A single leak of identity theft, financial, or medical records can damage both the client's reputation and the trust of customers. Therefore, ITES (Information Technology Enabled Services) companies must ensure strong data protection measures, such as secure infrastructure, restricted access, regular audits and stringent privacy policies, to maintain confidentiality as well as privacy to comply with local and international data protection laws like GDPR, or PDPL.

BPO - contextual relevance to the aviation big data ecosystem

In the aviation industry, BPO is widely adopted for functions such as customer service, ticketing, data analytics, loyalty program management and IT support. Airlines and airports increasingly rely on third party vendors to process huge amounts of passenger and operational data, including PII such as passport numbers, travel history, payment details and biometric data. Although outsourcing brings efficiency and cost savings, it significantly increases data privacy issues in the aviation ecosystem. For example, when a call center operator or data analytics firm in another jurisdiction handles passenger data, differences in privacy regulations (such as GDPR in the EU versus less stringent laws elsewhere) can lead to data misuse or unauthorized data access.

Occurrence probability (high - 0.6 to 0.8)

The likelihood of BPO related data privacy incidents in aviation is high and increasing, primarily due to the sector's heavy reliance on digital outsourcing and cross-border data exchange. Modern airline operations include numerous third-party integrations like reservation systems, customer experience analytics and payment processors, all of which require constant data flow between internal systems and external vendors. Certainly, each transfer point increases exposure. Moreover, BPO firms often serve multiple clients across industries, creating data related risks where aviation data might inadvertently mix with unrelated datasets. Given that many BPO centers are in countries with less robust cybersecurity infrastructure and lenient privacy laws, the probability of accidental leaks, insider misuse, or cyberattacks exploiting these intermediaries is significant.

Consequential impact – nonconformance BPO

The consequences of a BPO-Driven data privacy breach in aviation can be extremely severe, extending beyond financial losses to reputational damage, legal penalties and complete operational disruptions. A data compromise can weaken passenger trust, leading to revenue decline and loyalty loss, especially in a service industry where customer confidence is paramount. Additionally, airlines face stringent regulatory scrutiny under frameworks like GDPR and IATA's guidelines, where Non-Compliance can lead to multimillion-dollar fines. A data breach could also expose sensitive flight operations' data, potentially affecting aviation overall security. For instance, if outsourced analytics teams mishandle aircraft maintenance or crew scheduling data, it could lead to unauthorized insights into airline operations, posing not just privacy but also national security concerns for international carriers.

E-Governance (Electronic Governance)

E-Governance means using digital technology such as online portals to deliver government services (like tax filing or healthcare) and communicate with citizens as-and-when needed. Governments often gather massive amounts of personal data such as IDs, addresses, medical records, income detail and biometric information (such as fingerprints or facial data).

E-Governance privacy challenges are relatively unique, because it involves massive data collection, storage and sharing of citizens' personal information through big data ecosystem, that are mostly interconnected. In some cases, governments move faster toward digitalization than they establish strong data protection framework, leading to uncontrolled privacy rules about data ownership, consent management and data sharing. The integration of technologies like biometric IDs and facial recognition further raises surveillance concerns, as data collected for public service delivery may be repurposed for political or security monitoring. E-Governance systems allow excessive government monitoring of citizens' online behavior. Without strict privacy laws, the same systems built to serve people can be used to track, profile, or discriminate against them. Continuous surveillance undermines personal freedom and trust between citizens and the state. These issues make E-Governance privacy challenges distinctively complex, as governments must balance efficiency and transparency with the protection of personal rights in a Big Data Ecosystem. As per Sajid Momin et al., E-Governance faces unique privacy challenges and protecting this data from unauthorized access, misuse, or breaches is difficult due to the complexity of systems and multiple agencies involvement [3].

E-Governance - contextual relevance to the aviation big data ecosystem

E-Governance refers to the use of big data ecosystem by governments to improve competency, efficacy, transparency and accessibility of public services. In the aviation sector, this includes digital technology for flight approvals, border control, passenger data management, air traffic regulation and airport security coordination. These functions heavily rely on Big Data Ecosystems, such as PNR (Passenger Name Record), biometric verification and advanced analytics for safety and logistics. However, the integration of e-governance with such massive datasets raises serious privacy concerns. For instance, global systems like EU's Entry/Exit System (EES) or Pakistan civil aviation authority's passenger data integration involve real-time sharing of sensitive personal data between airlines and state databases. Without strict privacy controls and safeguards, these systems risk exposing confidential information to misuse, surveillance, or cyberattacks, highlighting why E-Governance in aviation is both a technological advancement and a privacy challenge.

Occurrence probability (moderate to high - 0.55 to 0.7)

The probability of privacy breaches within aviation e-governance systems is moderate to high, given the complexity of big data ecosystem and global interconnectivity. Airlines, airports and immigration authorities use multiple interoperable platforms, mostly involve third party service providers. This nature of heterogeneity and distributed data architecture increases the risk of data leakage, unauthorized access, or nonconformity with privacy laws such as KSA PDPL and GDPR standards. For example, incidents where passenger data was unintentionally exposed because of unsecured APIs or cloud storage have been reported by some international carriers. The aviation industry's greater dependency on continuous digital communication (for ticketing, tracking, immigration and security clearance) makes it highly vulnerable to privacy violations, not necessarily from malicious intent alone, but also from weak data privacy frameworks or technical lapses.

Consequential impact - regulatory shortfall in E-Governance

The consequences of privacy gaps in aviation e-governance can be severe and multidimensional. On an individual level, exposed personal data (e.g., biometric identifiers, travel history) can lead to identity theft or profiling. For airlines and aviation authorities, such breaches damage institutional credibility and may result in legal penalties, diplomatic tension and financial losses. As a case in point, the British Airways data breach exposed around 400,000 passengers'

financial details and led to impose, several million pounds in fines under GDPR and long-term trust erosion. In a national context, compromised aviation datasets can endanger state security, allowing the tracking of officials, intrusion of logistics systems, or manipulation of flight safety data. Therefore, privacy challenges in e-governance go beyond mere compliance, they directly influence operational resilience and public confidence in the aviation ecosystem.

Children's Data Privacy

Protecting children's data is extremely problematic as kids often don't comprehend the risks and criticalness of sharing their personal information online. Distinct apps and games may collect their name, age, location and even voice or video without legitimate consent. A typical example might be a mobile game, tracking a child's location to show nearby offers or ads, without informing the parents and their consent. This may trigger serious privacy risks, including exposure to strangers, targeted ads, or long-term misuse of their data. That's why special laws, tools and forceful privacy mechanisms are required to protect children's information and ensure parents' explicit consent.

Children's data - contextual relevance to the aviation big data ecosystem

In the aviation industry, children's data is increasingly collected through digital platforms such as online booking systems, frequent flyer programs, IFE (Inflight Entertainment) Systems, mobile apps and airport Wi-Fi services. Airlines and airports frequently gather sensitive information including a child's name, age, passport details, travel preferences and even biometric identifiers like facial recognition data, used at boarding gates. For instance, major international hubs in addition to Dubai International Airport have introduced automated boarding gates that use biometric verification, which sometimes involve children, traveling with families. This creates a unique privacy concern because children are considered a vulnerable group under most data protection laws, including GDPR and COPPA (Children's Online Privacy Protection Act) in the U.S. The aviation sector, consequently, faces a dual challenge: Ensuring seamless, data driven customer experiences while safeguarding children's personal data from misuse, profiling, or unauthorized sharing.

Occurrence probability (moderate to high - 0.55 to 0.7)

The probability of children's data being mishandled or exposed, in aviation systems is moderate to high, due to the large-scale digitalization of air travel and extensive global data sharing among airlines, airports, immigration authorities and third-party vendors. Booking portals and mobile applications are particularly vulnerable because they often rely on third party analytics, cookies and APIs that can inadvertently collect and share children's information. A realistic example is when family bookings are made through travel aggregators such as airline apps that request the child's date of birth to calculate ticket pricing, this information may then be stored or transmitted across multiple international channels. In addition, biometric systems at airports pose a rising risk of data leakage if state-of-the-art encryption, anonymization and other industry's best security practices are not in place. Considering the interconnected nature of aviation IT systems and the diversity of data processors involved, the likelihood of unintentional breaches or unauthorized data flows remains significant.

Consequential impact - compromised children's data

The negative consequences of exposed children's data in the aviation sector can be severe and long lasting. A data breach could result in identity theft, unauthorized profiling, or exposure of travel histories that put children at physical or psychological risk. Moreover, reputational damage to airlines and airports can weaken public trust and trigger regulatory fines and penalties. For example, if an airline's system were to expose the personal details of child passengers, it could face fines under GDPR and beyond financial loss, there is also an ethical impact. Children may not understand or consent to how their information is used and hence their data could be leveraged for targeted advertisements, behavioral analytics and exposed to strangers. Such abuse undermines public confidence in digital aviation services and raises moral questions about informed consent, especially when children's digital identities are formed without adequate protection.

Automation and Orchestration

Automation and orchestration are at the core of modern Big Data ecosystems, enabling organizations to process, analyze and act on vast quantities of data with remarkable speed and minimal human intervention. Automation executes repetitive, rule-based tasks such as data collection, labeling, or real time analytics, by means of tools and technologies such as APIs, Python scripts, Power Automate, Function-As-Service i.e. AWS Lambda and OCI Function. Orchestration coordinates, organizes and manages automated workflows across multiple systems, clouds and tools using distinct enablement methods such as Apache airflow, Kubernetes, OCI resource manager, ansible tower and azure logic apps. altogether, they enhance overall efficiency and usefulness. However, they also introduce and magnify privacy risks when handling personal or sensitive data. For instance, in aviation, orchestrated data flows may integrate passenger information from booking platforms, biometric scanners and security systems to create unified traveler profiles.

While this improves operational efficiency and passenger experience, a single misconfigured automation rule or poorly secured orchestration link can expose entire datasets, revealing travel histories, health details, or identity markers. Similarly, in healthcare or finance, automated decision-making pipelines can process millions of records in seconds; a faulty algorithm or unauthorized orchestration of third-party APIs can lead to widespread privacy breaches. Moreover, regulations like the PDPL, GDPR and CCPA demand contextual interpretation such as enforcing data minimization or the "right to be forgotten", which automated systems often cannot fully apply, resulting in compliance gaps and potential legal exposure. Another significant challenge is ensuring auditability, accountability and control across these complex orchestrated processes and automated networks. Each automated and orchestrated process generates enormous volumes of logs, transactions and algorithmic outputs, making it difficult to trace how data was acquired, stored, altered, or shared. For example, in aviation or smart city management, orchestrated AI systems that continuously update passenger or citizen risk scores may produce thousands of automated privacies impacting decisions every hour decisions that are nearly impossible to review manually.

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

Automation and orchestration - contextual relevance to the aviation big data ecosystem

In the aviation industry, automation and orchestration have become indispensable for managing the vast and complex big data ecosystems that fortify modern air travel. Airlines, airports and regulatory authorities depend heavily on interconnected data systems for functions such as flight scheduling, predictive maintenance, air traffic coordination and passenger experience personalization. Automation ensures that repetitive, data intensive tasks like biometric verification, ticket validation and baggage tracking, are executed swiftly and accurately, while orchestration coordinates the seamless flow of this data across multiple subsystems, including reservation platforms, security databases and cloud-based analytics engines. Passengers' information, which may include biometrics, travel histories, payment credentials and even health data (e.g., vaccination or accessibility requirements), passes through multiple automated and orchestrated systems in real time. A single failure in terms of Automation and Orchestration, like a misconfigured data pipeline or unauthorized orchestration link (between automated tasks) can expose sensitive information to unintended systems. For instance, when orchestrated analytics combine flight manifests with immigration and security datasets, weak access controls or faulty synchronization can result in data oversharing across jurisdictions. Furthermore, compliance with aviation specific data privacy frameworks and international regulations such as GDPR, CCPA and ICAO data protection guidelines is complicated by the distributed and dynamic nature of orchestrated environments.

Occurrence probability (higher - 0.79)

The probability of privacy risks originating from automation and orchestration is higher in modern big data ecosystems due to the increasing reliance on autonomous and interconnected systems. Studies and industry reports consistently show that misconfigured automation scripts, insecure orchestration links and third-party integrations are among the most frequent causes of large-scale data breaches. Automation related risks (e.g., algorithmic misclassification, unmonitored data sharing) have a probability rating around 0.6-0.8, since such processes occur thousands of times daily with limited oversight. Orchestration related risks (e.g., data leakage between orchestrated systems or cloud platforms) often range between 0.5-0.7, especially where multi-vendor or cross border data flows exist. Overall, the combined occurrence probability of automation and orchestration driven privacy incidents is classified as "upper high (0.79 likelihood)", meaning such events are likely to occur several times within a year in large organizations unless mitigated through suitable data engineering, robust data privacy framework and continuous compliance monitoring.

Consequential impact – unregulated automation and orchestration

Failures or breaches because of automation or orchestration in the aviation sector can have far reaching and awful consequences that extend well beyond technical disruption. The most immediate negative impact is the loss of passenger privacy and trust, as exposed personal data such as travel itineraries, biometric identifiers, or payment information can be exploited for identity theft, targeted cyberattacks, or surveillance misuse. When automated pipelines share or process data without adequate orchestration control, breaches may propagate rapidly across interconnected systems, extending the scale of exposure. This can severely damage an airline's reputation and weaken public confidence in digital aviation systems that rely on passengers' willingness to share personal data for safety and convenience. Operationally, privacy vulnerability can trigger system downtime, flight delays, or service interruptions, especially when compromised automation scripts or orchestrated workflows are disabled for investigation. On a regulatory level, noncompliance with data protection laws such as PDPL, GDPR and CCPA can lead to substantial financial penalties often amounting to millions of dollars as well as legal restrictions on data transfers. Furthermore, the aviation sector's dependence on cross border data flows means a single breach can invoke multiple jurisdictions, complicating legal accountability and response coordination. Beyond economic loss, such incidents may expose critical operational data (e.g., flight paths or crew schedules), pose national security and safety risks if exploit by malicious actors.

Balance between Innovation and Data Privacy

One of the great challenges is the game between data profitability and data privacy. According to Mulder and Vellinga (2021) [9], "in principle the processing of sensitive data is prohibited, unless one of the exemptions mentioned in Article 9 (2) GDPR applies". Data financial viability, individuals' autonomy and data privacy are imperative. Evidence suggests that "In a world where data drives technological advancements, the tension between innovation and privacy becomes more pronounced. The challenge lies in creating a balance that preserves the benefits of data utilization while protecting individual rights" [6]. Certainly, it is really very hard to balance usefulness (e.g., location patterns, maps, delivery etc.) with privacy. Using advanced hacking techniques, attackers can reconstruct an individual's identity by piecing together multiple small fragments of information. In principle, "The greater the privacy, the lesser the utility and equally, the greater the utility, the lesser the privacy." As per Goel et al., "The tradeoff between preserving privacy and the benefits gained from data utilization is to be considered seriously" [12].

Innovation and data privacy - contextual relevance to the aviation big data ecosystem

In the aviation industry, the balance between innovation and data privacy has become a focal challenge due to the growing reliance on digital technologies such as AI Powered predictive maintenance, passenger analytics and biometric security systems. Airlines and airports increasingly use big data to enhance operational efficacy, safety and enhance customer experience. For example, by predicting flight delays, optimizing fuel consumption, or personalizing passenger services. However, these innovations rely on collecting vast amounts of sensitive information, including biometric identifiers (facial recognition), geolocation and travel history. When systems like IATA's One ID (IATA-driven digital identity program) or digital health passports are used, data privacy concerns amplify as passenger information is shared across multiple international stakeholders. Thus, innovation in aviation cannot progress sustainably without strong safeguards ensuring that privacy and data protection are integrated into every stage of data management lifecycle.

Occurrence probability (high - 0.6 to 0.8)

The potential privacy risk, resulting from emerging technological adoption, in the aviation sector is high, given the sector's increasing digitization and global data flows. The probability is elevated by the extensive use of cloud-based storage, third party analytics vendors

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

and interconnected systems between airlines, airports and government agencies. For instance, when a major airline introduces an AI-based recommendation system for in-flight services, data might be processed by external partners or stored in jurisdictions with weaker data protection laws. Even with robust regulatory frameworks like EU-GDPR or CCPA, enforcement inconsistencies across countries create loopholes. Historical incidents, such as the British airways data breach of 2018, which exposed personal data of over 400,000 passengers, highlight how easily privacy violations can occur, especially when innovative systems outpace existing privacy and security compliance mechanisms.

Consequential impact – imbalanced data profitability and data privacy

The negative impact of failing to balance innovation and privacy in aviation can be extremely severe and multidimensional. From a regulatory standpoint, airlines risk heavy fines, legal penalties and loss of operating licenses. Economically, a privacy breach undermines customer confidence and brand reputation, leading to decreased overall ticket sales and partnerships. On an operational level, compromised systems may disrupt flight operations or national security coordination. For example, if biometric boarding data were mishandled or leaked, it could not only violate individual rights but also destabilize airport security protocols and international trust among regulators. Beyond immediate losses, such breaches can also delay the adoption of beneficial technologies like AI-Powered passenger verification or digital identity platforms, as public uncertainty rises toward innovation, perceived as unwelcome.

Regulatory Disparities Across Border

According to Ijaiya, the GDPR, introduced in 2018, is often lauded as the gold standard for data protection, with its emphasis on user consent, data portability and the right to be forgotten [6]. In contrast, the CCPA focuses on giving consumers control over their data by enabling them to opt out of data sales and request access to collected information. Harmonized policies ensure consistency in addressing privacy issues while reducing compliance challenges for multinational corporations. However, challenges persist, such as geopolitical tensions and differing priorities among nations. For instance, while the European Union prioritizes stringent data protection measures through GDPR, other regions focus on fostering innovation and economic growth, sometimes by compromising privacy control. Bridging these gaps requires sustained dialogue and collaboration among global stakeholders. Regarding the considerable variation in laws, McCausland stated that “These laws vary, creating challenges for privacy and data protection, especially when data and personal information crosses borders. In our globalized world, the G7 and others now recognize the need for more international cooperation to prevent fragmentation and promote interoperability and certainty for consumers and businesses”. McCausland, also highlighted the challenges as follows [18]:

- Data protection laws exist globally but are inconsistent.
- Multiple data protection laws may apply simultaneously.
- Data protection laws are increasingly conflicting with other laws.
- Data protection laws create barriers to cross-border data flows.
- Data localization requirements are unworkable.

The legal requirements for data collection, processing and cross-border transfer vary between different jurisdictions which presents challenges to effective operation of civil aviation and delivery of services as per customers’ expectations. Where EU-GDPR applies or where the local model looks to this standard of law, the “controller” is required to have a legal basis in place, especially when cross-border data sharing is going-on, otherwise the processing will be considered illegitimate. Conversely, in the USA, the California Consumer Privacy Act does not require the particular “legal basis” for the initial collection of data to be specified, only to provide transparency on the “business or commercial purpose” for collecting the data. According to Mulder and Vellinga, “both the public interest in road safety and the personal right to data protection can co-exist” [9].

The territorial scope of Big Data operations highlights one of the most complex privacy challenges in today’s interconnected digital ecosystem, as mentioned by Voss, [7]. Big Data is inherently global and borderless, where information is continuously collected, stored and processed across multiple jurisdictions, each governed by distinct and often conflicting privacy regulations. This global dispersion creates serious compliance and ethical dilemmas. A company’s data practices may fully adhere to one nation’s legal standards while simultaneously breaching another’s.

An additional pressing challenge is “privacy laws like the GDPR and CPRA are not written prescriptively, to tell you exactly what to do, but are principle based.” [5]. These laws and regulations set broad principles, expected outcomes and ultimate goals (such as Fairness, Transparency, Accountability and Purpose Limitation), instead of providing sequential steps, rules, or technical checklists on what exactly to do. This could trigger the following:

- **Ambiguity in implementation:** A small airline may not know what counts as “Adequate Protection” for passenger data.
- **Compliance uncertainty:** Organizations fear regulatory fines and penalties because “Adequate”, "Appropriate" or “Reasonable” is not defined. GDPR requires “appropriate security measures”, what’s enough is not clear.
- **Resource inefficiency:** Organizations may overcompensate or undercompensate, wasting resources or taking risks. Unnecessarily investing in expensive encryption when simpler controls may suffice or doing too little and being noncompliant.
- **Variability across organizations:** Different interpretations lead to inconsistent privacy protection. Two airlines may handle passenger data very differently, yet both claim compliance.
- **Difficulty auditing:** Auditors or regulators must judge if the company’s approach meets the principles. For example, DPIAs may be “technically done” but still not actually protective.

To address all the highlighted challenges relating to regulatory disparities, organizations can adopt a range of mitigation strategies, which may include (but not limited to):

- Implementing data localization policies such as PDPL, ensuring sensitive information remains within national borders.
- Establishing SCC (Standard Contractual Clauses) as emphasize by Williamson and Prybutok 14]. “Addressing the legal and regulatory challenges involves drafting

comprehensive contracts that clearly define the rights and obligations of all stakeholders".

- Maintaining strong BCRs (Legal freedom + Ethical trust), as recommended by McCausland [18]. "In the alternative, BCRs are another available mechanism, but these only operate as between companies within the same group and cannot be used as a contract with third party entities". BCR allows global data mobility (within same group) without compromising privacy integrity.
- Performing proper DPIA (Data Protection Impact Assessments) by involving multidisciplinary teams (like Cybersecurity, IT, Business) before global expansion. Strong DPIA ensures trust, compliance and operational safety, not just paperwork.
- Aligning with international privacy frameworks such as the GDPR's adequacy decision mechanism and the APEC (Asia-Pacific Economic Cooperation) Cross-Border Privacy Rules (CBPR) system, which help promote interoperability and ensure a more consistent level of protection for individuals' personal data worldwide.

Regulatory disparities - contextual relevance to the aviation big data ecosystem

In the aviation industry, the challenge of regulatory disparities across borders is particularly pronounced due to its global operations. Airlines, airports and aviation service providers constantly exchange vast volumes of passenger data across jurisdictions including booking details, biometric information, flight records and security data. However, data protection laws vary significantly from one region to another. For instance, countries in Asia region and the Middle East have emerging yet inconsistent privacy frameworks. This regulatory fragmentation complicates data handling for multinational airlines such as Emirates, Saudi Arabian Airlines, PIA, Lufthansa, or Singapore Airlines, which must adapt their systems to comply with multiple overlapping and sometimes conflicting, privacy requirements.

Occurrence probability (high - 0.6 to 0.8)

The probability of encountering regulatory disparities in the aviation sector is high, given the industry's dependence on cross border collaboration and digital ecosystems. Airlines routinely use third party platforms for reservations, loyalty programs and biometric check-ins, often relying on cloud services hosted in different regions. The frequent transfer of passengers' data between jurisdictions (e.g., from an EU-based travel agency to an airline's servers in the Middle East or the U.S.) increases the possibility of conflicting compliance obligations. Moreover, with the rise of AI driven predictive maintenance, customer analytics and smart airport systems, the collection and analysis of big data have become even more complex. These operations often require integrating datasets from multiple sources, each governed by local data privacy laws. As aviation companies expand globally and adopt digital solutions, the probability that regulatory inconsistencies will result in unavoidable compliance challenges remains high.

Consequential impact - variation in regulatory frameworks

The impact of these regulatory disparities and variation can be severe and multidimensional. From a legal perspective, noncompliance with local data protection laws may lead to substantial fines and heavy penalties. For instance, GDPR violations can result in fines of up to €20 million or 4% of annual global turnover. Likewise, airlines may face disruptions if certain data flows are restricted by strict jurisdictions, affecting services such as passenger processing, flight coordination, or emergency response systems. Moreover, there might be reputational risks such as passengers losing trust in how their personal data is handled, which can lead to brand damage and reduced customer loyalty. Furthermore, inconsistent regulations can hinder the seamless implementation of global safety and health initiatives, such as the IATA Travel Pass, which depends on reliable and legally compliant data sharing.

Opaque Consent Management

It is a major big data privacy challenge because it allows organizations to collect vast amounts of personal information without users fully understanding what they agree to. According to Ijaiya [6]. "Consent and data ownership are additional areas of contention in AI Driven data analytics. Many users remain unaware of how their data is collected, processed and used, often due to opaque consent mechanisms buried in lengthy terms and conditions. This lack of informed consent erodes trust and raises ethical questions about the legitimacy of data driven decisions". A shopping application might ask for permission to "improve your experience" without drawing attention to that data will also be sold to advertisers and marketing agencies. Health tracker may bundle consent in a single click "Accept All" button, making it unclear that sensitive medical data will be shared with third parties. These practices weaken informed consent, enable hidden large-scale data sharing and reduce trust, especially when data is aggregated into Big Datasets for profiling and prescriptive analytics.

Consent management - contextual relevance to the aviation big data ecosystem

In the aviation industry, opaque consent management represents a critical challenge due to the massive volumes of passenger and operational data collected across multiple touchpoints. Airlines, airports and travel service providers gather information through ticket bookings, mobile apps, loyalty programs, biometric boarding and in-flight services like Wi-Fi usage. Often, passengers consent to broad, unclear or ambiguous terms without fully understanding how their data will be processed and shared with third party analytics firms, marketing agencies, or border control systems. For example, a passenger agreeing to "enhanced travel experience" terms may unknowingly allow cross sharing of their geolocation and browsing data between the airline's partners. This lack of transparency compromises informed and explicit consent, exposing aviation companies to regulatory scrutiny under GDPR, CCPA, PDPL and similar frameworks, ultimately minimizing public trust in digital travel ecosystems.

Occurrence probability (high - 0.6 to 0.8)

The probability of opaque consent practices in aviation is high, primarily because digitalization and data driven personalization have become standard competitive tools. Airlines rapidly move toward predictive maintenance systems, dynamic pricing algorithms and customer behavior analytics that integrate data from distinct multi-modal sources. In such complex data flows, consent obtained at one stage (e.g., mobile check-in) may not cover subsequent data processing (e.g., AI based marketing or data sharing with security agencies and third parties). Moreover, consent exhaustion, where

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

users routinely accept terms without reading them, amplifies the opacity. For instance, during online ticket booking, passengers must approve long, multi-page privacy statements, which often include data sharing clauses. Given the global nature of aviation, differing regional privacy regulations make consistent consent management even more difficult, raising the likelihood of inadvertent noncompliance.

Consequential impact – obscure and ambiguous consent

The negative consequences of opaque consent management in aviation are multi-layered, affecting both organizational and societal dimensions. Legally, airlines may face considerable fines and sanctions for violating data protection laws such as CCPA, PDPL, or GDPR. For instance, if an airline shares passenger data with third party analytics firms without explicit consent, it could be held liable for unauthorized processing. Operationally, unlawful data processing can lead to reputational damage and loss of customer loyalty, especially in a safety critical industry like aviation, where "Trust is the Top Priority". Additionally, there are security effects, unclear data flows might expose sensitive travel or biometric detail to cyberattacks or misuse by malicious actors, potentially compromising both individual privacy and national security interests.

Enterprise-Level Obstacles to Privacy Governance

Organizations face multi-dimensional unique series of technical, human, structural and strategic challenges in ensuring big data privacy and protection within an emerging and interconnected digital ecosystem. The exponential growth of data volume, variety and velocity has drastically increased privacy risks, making it very difficult to control access, ensure lawful processing and maintain transparency across complex data flows. Internal barriers such as limited resources, fragmented privacy responsibilities and the absence of measurable privacy performance metrics often create unbearable challenges from organizational readiness perspective. Weak collaboration between IT, cybersecurity, data Governance and compliance teams creates gaps in accountability, whereas complacency due to a perceived low risk of penalties leads to insufficient investment in data privacy related projects. Moreover, organizations struggle to adapt to rapidly evolving privacy regulations, such as GDPR, CCPA and other regional frameworks like PDPL resulting in inconsistent data protection and privacy practices. Furthermore, On the operational side, behavioral and insider risks, including employee negligence or deliberate data misuse, further threaten data confidentiality and trust. Strategically, many organizations still perceive big data privacy as a regulatory requirement rather than a business enabler, neglecting its role in driving customer confidence and sustainable innovation. This fragmented approach widens the gap between the pace of big data expansion and the organization's ability to safeguard personal and sensitive information. To enhance further understanding, the subsequent four key enterprise level challenges are crucial to shape big data privacy practices, providing clarity on how each one influences an organization's privacy posture.

Organizational structure and collaboration

As stated by Kajunju and Hamilton, organization hierarchy misalignment, deciding which part of the organization the privacy team should belong to and reporting structure might be great challenges for privacy team and contribute substantially to their overall productivity [5]. Secondly, leadership and departments are disconnected and privacy often sits too low in the organizational hierarchy to influence decision-making. Moreover, fragmented privacy responsibilities reflect unclear ownership and privacy related tasks are spread across departments with little coordination. Furthermore, inconsistent ontology and taxonomy show lack of shared language, vocabulary and framework, eventually obscures classifying and managing personal data consistency.

Enterprise perception and strategic value

Due to intangible nature of privacy value, privacy may be viewed as slowing down or limiting business opportunities and generally it is really a big challenge to quantify ROI (Return-On-Investment) of data privacy. Protecting privacy brings benefits that are hard to measure, making it difficult to justify sustained investment. Secondly, privacy is also seen as a compliance burden. Without clear business value, privacy initiatives are deprioritized against more visible objectives. Kajunju and Hamilton underlined that "the privacy team should take steps to ensure the business understands the importance of privacy to the organization and provide metrics to measure its performance" and privacy team should also "tie their work to key business objectives" and "establish privacy as a business enabler" [5].

Operational and behavioral risks

Organization insiders considered a major Big Data privacy challenge because employees often have legitimate access to sensitive information such as customer records, financial data and company secrets. Unlike external hackers, insiders are already trusted within the system, which makes their actions harder to detect. Ramya et al., highlighted that "employee theft is prevalent not only in big tech companies but also in startups" [16]. Privileged inner-circle members' access means that one dishonest employee can misuse or steal vast amounts of data without raising immediate suspicion.

Regulatory and external drivers

Alongside the earlier identified enterprise level challenges, organizations continue to experience other critical internal and strategic challenges, which are summarized as follows:

- **Complacency due to perceived low risk of fines:** Organizations may underinvest in privacy because penalties feel remote or unlikely, as stated by Kajunju and Hamilton, "the potential for regulatory penalties is viewed by businesses as unlikely" [5].
- **Difficulty adapting to evolving privacy regulations:** Constantly changing laws across jurisdictions create uncertainty and compliance fatigue.
- **Lacking established metrics in privacy programs:** The absence and lack of established metrics to measure the success of a privacy program, makes it difficult to explain the value and importance of privacy [5]. Organizations are unable to define what success looks like, making it difficult to track effectiveness, identify weaknesses, or demonstrate accountability. Without measurable benchmarks, programs risk becoming reactive rather than proactive, as they cannot provide early warning signals for potential risks or evaluate whether initiatives like training, compliance checks, or incident response are working. This lack of clarity also undermines trust with regulators, executives and customers who increasingly expect evidence of progress.

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

- **Limited resource:** Resources are limited, while the scope of work is extensive due to the wide variety and sheer number of regulations and laws. “Privacy team is small Privacy teams are generally very small parts of the organizations they serve” [5].

Enterprise-level data privacy obstacles - contextual relevance to the aviation big data ecosystem

The aviation industry manages gigantic amount of sensitive personal data, from passenger identities, biometrics, travel histories, payment details, to aircraft telemetry shared across global systems. This makes privacy governance not only a compliance obligation but also a cornerstone of passenger trust and international cooperation. However, enterprise-level challenges such as fragmented privacy responsibilities and hierarchical misalignment are prevalent in aviation sector, where operations span multiple subsidiaries, airports and international regulatory frameworks. For instance, one division might handle booking systems under EU-GDPR, while another manages biometric boarding under FAA (Federal Aviation Administration US) frameworks. Inconsistent privacy taxonomies across these entities create data silos, making unified governance challenging. Additionally, privacy is often viewed as a compliance burden rather than a strategic asset, limiting innovation in Data Driven Personalization or AI-based maintenance analytics. To overcome these challenges, enterprises must embed privacy governance into every stage of data management lifecycle, establish clear accountability frameworks and promote a privacy-aware culture, ensuring that data privacy and protection evolve in parallel with big data innovation, compliance and ethical responsibility.

Occurrence probability (high - 0.6 to 0.8)

The probability of encountering such governance hindrances in aviation is high, because of the sector’s complex organizational hierarchies, global regulatory exposure and reliance on third party data exchanges. Airlines often outsource critical IT and customer data operations to vendors, increasing fragmentation and mitigating accountability. For example, a leading airline might have separate systems for loyalty programs, ticketing and airport operations, each managed by different business units or external partners. When privacy oversight is distributed in this manner, employee behavior risks (like unauthorized access or data theft) and misalignment between compliance and IT teams are common. Moreover, the intangible nature of privacy value, unlike easily measurable KPIs such as fuel efficiency, means that decision-makers may deprioritize privacy investments until a breach or fine occurs.

Consequential impact - organizational data privacy failures

Failure to effectively manage internal data privacy issues has a very high negative impact on the organization. It weakens ethical integrity by fostering fragmented accountability and weak privacy culture, signaling negligence in valuing personal information and transparency. This ethical issue cascades into high reputational damage, as stakeholders, customers and regulators perceive the organization as unreliable and noncompliant. Inconsistent privacy frameworks, ontologies and taxonomies intensify governance confusion, amplifying visibility of internal weaknesses. Operational lapses and insider risks further increase compliance exposure, while complacency toward regulatory obligations results in systemic noncompliance and severe regulatory vulnerability. Collectively, these effects lead to significant ethical, reputational and compliance degradation, critically discourage organizational credibility, customer trust and sustainable growth. Moreover, fragmented privacy structures increase the likelihood of data breaches or unauthorized cross border transfers, exposing airlines to substantial penalties under regulations like the PDPL, GDPR, CCPA, or PDPA.

Data Analytics

In the era of big data, both data analysis and data analytics cause significant challenges to individual privacy. These processes involve collecting, processing and interpreting vast volumes of personal, behavioral and transactional information from diversified sources such as social media, healthcare systems, financial networks and IoT devices. When these heterogeneous datasets are combined, they often generate detailed digital profiles that enable the re-identification of individuals, even when the original data was anonymized.

For instance, in healthcare analysis, cross-referencing anonymized hospital records with demographic information can unintentionally disclose patient identities. Similarly, retail corporations analyzing customer activity across distinct online platforms may infer deeply private attributes such as income level, health status, or political preferences. These risks are combined by the fact that most of the users have limited awareness about how their data is collected and reused, resulting in violations of informed and explicit consent. Furthermore, the large-scale storage and transfer of data increase the probability of breaches and unauthorized access. Each transfer, storage point, or access permission becomes a potential entry point for cybercriminals and unauthorized users. Indeed, data analysis provides insights and data analytics drives innovation and efficiency, but both processes, at the same time, complicate the protection of personal privacy (Figure 7).

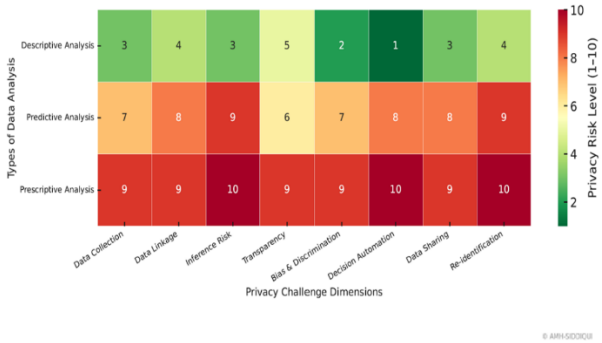

Figure 7: Big Data privacy challenges dimensions - across analysis types.

Descriptive analysis

Descriptive analysis serves as the foundation of data analytics by summarizing large datasets to identify trends and patterns such as average income levels, age distributions, or prevalent medical conditions. However, even when data is aggregated, privacy concerns persist. For example, if a hospital releases statistics showing that 78% of patients in a small community suffer from a specific disease, individuals in that locality could still be indirectly identified. The same issue might happen when consumer behavior is described using purchase histories or location-based data; these summaries, when associated with external datasets, can reveal personal habits or

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

lifestyles. Therefore, although descriptive analysis enhances, let's say, our understanding of populations, it carries an inherent risk of indirect identification, profiling and potential misuse of sensitive information.

Predictive analysis

Based on descriptive insights, predictive analysis utilizes historical data to forecast future outcomes such as disease risk, credit default, or consumer purchasing behavior. While predictive models provide valuable foresight, they also introduce serious privacy challenges because they depend heavily on sensitive data. Secondly, predictions can expose personal attributes that most of the individuals never consented to share. Example involves a retail company that inferred a teenager's pregnancy through shopping patterns before her family was aware, prompting targeted advertisements. Likewise, financial institutions may use predictive algorithms to classify individuals as high-risk borrowers, based on social media activity or geographic location, which can lead to discrimination and ethical controversies. Consequently, predictive analysis, though powerful, often blurs the line between legitimate insight and annoying surveillance.

Prescriptive analysis

Prescriptive analysis extends predictive analytics by recommending specific actions or decisions through advanced algorithms and machine learning. It is used in domains such as healthcare, finance, aviation and marketing to suggest optimal treatments, credit limits, dynamic flights rebooking, or personalized advertisements. However, these recommendations depend on the extensive collection and integration of personal data from multiple sources including medical records, online behavior, sentiments and social media activity. Similarly, an online retailer might target users with emotionally tailored advertisements based on inferred stress levels or financial strain, crossing ethical and privacy boundaries. Hence, while prescriptive analysis aims to optimize decision-making, it can ultimately lead to manipulation and data misuse by allowing algorithms to make choices for individuals without their knowledge and explicit consent.

Data analysis - contextual relevance to the aviation big data ecosystem

In the aviation industry, data analysis including descriptive, predictive and prescriptive approaches plays a crucial role in overall operational efficiency, safety and passenger experience. Airlines and airport authorities rely on big datasets from aircraft sensors, maintenance logs, weather systems, passenger bookings and flight operations. Descriptive analytics help summarize past events, such as fuel consumption or flight delays. Predictive analytics use machine learning to forecast events like equipment failures or passenger no shows, while prescriptive analytics recommend optimal actions such as maintenance schedules or route adjustments. However, the integration of these analytical methods into aviation systems introduces significant privacy challenges. For instance, predictive maintenance tools that analyze sensor and pilot behavior data may inadvertently expose sensitive operational details or personal information about crew and passengers, if data privacy framework is weak. Similarly, analyzing biometric data for security screening or passenger movement tracking raises ethical and privacy concerns under data protection regulations such as PDPL, GDPR or ICAO's cybersecurity frameworks.

Occurrence probability (high - 0.6 to 0.8)

The probability of data privacy risks occurring within aviation data analytics is high, primarily due to the volume, variety and velocity of data collected in this sector. Airlines and airports manage terabytes-to-petabytes of personal and operational data, from booking platforms, loyalty programs and mobile apps to in flight Wi-Fi and maintenance systems. The likelihood of data breach, misuse, or unintended data sharing increases when multiple vendors and third-party analytics providers (which are inherent in aviation) are involved. For example, predictive models trained using passenger travel histories and payment data can unintentionally leak PII (Personally Identifiable Information) if anonymization or encryption is insufficient. As aviation increasingly adopts AI driven analytics for forecasting and optimization, the probability of privacy compromise will remain significant unless strong data governance mechanisms are enforced.

Consequential impact – unethical and unregulated data analysis practices

The consequences of privacy violations in aviation data analytics are extraordinary harmful and destructive, affecting trust, compliance, safety and may cause operations' suspension. Breaches can lead to massive reputational damage, regulatory fines and loss of customer confidence. Under GDPR, for example, data controllers face penalties of up to 4% of annual turnover for privacy noncompliance. For airlines, compromised passenger data could result in identity theft, unauthorized tracking, or even national security threats if sensitive travel information falls into the wrong hands. Moreover, exposure of operational or flight data could affect competitive advantage or compromise aviation safety systems that rely on secure communication channels. Predictive or prescriptive analytics errors could also have real world consequences such as inaccurate maintenance predictions leading to flight cancellations or safety hazards. Beyond financial loss, such events can destroy public trust in AI based aviation technologies, discouraging further innovation (Figure 8).

Figure 8: Proposed remedial framework for big data privacy.

Proposed Comprehensive Remedial Framework for Big Data Privacy

Addressing the multi-dimensional data privacy inbound and egressing challenges arising from distinct innovations and technologies (AI, ML, LLM, NLP, automation, orchestration etc.), e-governance, regulations and large-scale data ecosystems in the aviation sector requires a multi-layered, forward thinking and robust governance-driven approach.

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

As depicted in figure 8, the illustration shows suggested framework which provides a holistic strategy encompassing technical, organizational, regulatory and ethical measures, ensuring that the efficiency of automation and advanced analytics is balanced with strong data protection, privacy and public trust.

Privacy-by-design and privacy-by-default integration

Aviation systems must embed strong privacy controls into every architectural layer from passengers' booking to predictive and prescriptive maintenance. Data minimization, anonymization, enhanced encryption and purpose limitation should be designed into workflows from inception, not added reactively. Fundamental measures may include (but not limited to):

- Integrating strong privacy preserving techniques such as differential privacy and federated learning to enable analytics without data exposure.
- Using tokenization and zero-trust architecture for all automated and orchestrated workflows.
- Ensuring data classification and minimization under a centralized privacy ontology to reduce unnecessary retention or duplication. This approach will ensure that organizations collect only the data they truly need, keep it as long as necessary and only use this data for the mentioned purposes (Figure 9).

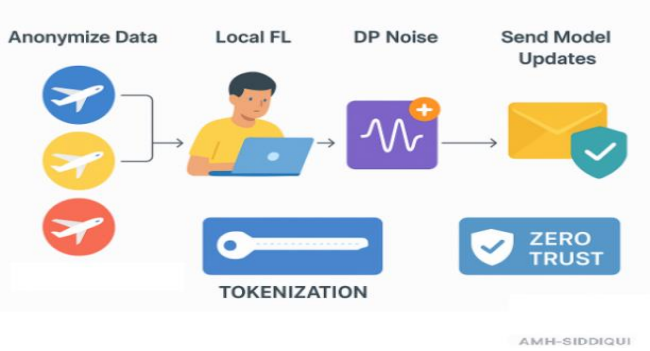

Figure 9: Privacy preserving technique pipeline.

Human-in-the-loop (Oversight)

Automation cannot and should never be fully replaced by human supervision, control and accountability. Systems performing risk scoring, identity verification, or passenger screening must retain human validation checkpoints. This prevents algorithmic bias and ensures lawful decision-making. Airlines must institute ethics review committees to audit automated processes and evaluate data handling fairness before any deployment.

Secure orchestration and real-time data integrity

A secure orchestration framework must monitor every automated data flow, considering the subsequent safeguards:

- End-to-end enhanced encryption (at rest, in motion and in use).
- Zero-trust authentication between all orchestrated systems and cloud nodes.
- Immutable data lineage tracking using blockchain to trace access, origin and usage.
- Routine integrity checks, balancing human supervision and AI-Powered anomaly detection to prevent data manipulation or leakage.

Immutable automated auditing and continuous monitoring

AI-enabled auditing tools should continuously scan automated workflows to detect anomalies, policy violations, or unauthorized transfers. Immutable audit logs aligned with PDPL, GDPR, CCPA and ICAO guidelines ensure traceability. Real-time alerts and automated isolation mechanisms reduce incident response time. Continuous compliance dashboards provide visibility for regulators and privacy officers. Williamson and Prybutok, stated that "despite these challenges, blockchain technology promises to enhance certain aspects of GDPR compliance [14]. For instance, blockchain can serve as a robust tool for data governance, offering new ways to manage and distribute data without relying on central authority. This decentralized approach can increase transparency in data transactions and streamline the process of data sharing. Blockchain can potentially empower individuals with greater control over their data, which aligns with the GDPR's objectives. For example, blockchain can facilitate the exercise of rights such as access to personal data (Article 15 GDPR) and data portability (Article 20 GDPR). These provisions give individuals a say in how their data is used and quickly transfer it from one service provider to another. By leveraging blockchain, data subjects could monitor how their data are being used, ensuring adherence to GDPR principles like purpose limitation and quickly identifying any breaches or misuse of their data".

Centralized privacy governance and accountability

Aviation enterprises must implement a centralized privacy governance model led by a CPO (Chief Privacy Officer) empowered across legal, IT, cybersecurity and operations divisions. Major elements may include:

- Standardized taxonomy for personal and operational data classification.
- Privacy Steering Committees to ensure that each digital transformation (e.g., biometric boarding, e-ticketing) includes adequate DPIA and proper risk assessments.
- Framing privacy as a strategic differentiator that promotes travel services (flights, bookings, loyalty programs, etc.) that are built around strong data privacy and security and use that protection trust as a selling point for the organization.

Transparent and granular consent management

In order to address opaque consent management, organizations advised to implement the following.

- Introduce layered consent interfaces that allow passengers to easily choose specific data uses (e.g., marketing, analytics, third-party sharing).
- Blockchain-Based Consent Management and Tracking are strongly advisable as Blockchain ensures immutable auditing for when, where and how consent was granted.
- Conduct regular PIAs (Privacy Impact Assessments) and data-flow mapping to detect illegitimate or unethical data processing.
- Align with IATA One ID and privacy by design standards for global consistency. IATA One ID is an emerging domain, a transformative shift in aviation passenger

processing, from documents at airport counters to digital identity + biometrics, pre-travel authorizations.
- Run passenger education campaigns to reinforce informed consent and trust.

Harmonizing regulatory disparities and global compliance

As discussed earlier, given the global nature of aviation industry, regulatory fragmentation poses remarkable data privacy risks. To minimize and mitigate these risks, organizations should:

- Adopt the "Highest Common Denominator" principle e.g., applying GDPR level safeguards globally.
- Conduct DPIAs for all international data transfers.
- Appoint regional DPOs for localized compliance.
- Utilize BCRs and SCCs for lawful cross-border data flow.
- Engage with regulators and international bodies such as ICAO and IATA to build interoperable privacy frameworks.

Data quality, accuracy and integrity protection

Deficiency in data quality, challenges safety, compliance and analytics, consequently impacting overall productivity, performance, trust and organizational reputation. To lessen the risks and fortify analytics capabilities, organizations must ensure regulatory compliance and incorporate the subsequent suggestions.

- Implement AI enabled anomaly detection and redundant validation mechanisms to ensure accurate flights, maintenance and passengers' data.
- Adopt ISO 8000 (data quality) and blockchain based audit trails to enhance traceability. ISO 8000 refers to the international standard for data quality and master data, developed and published by the international organization for standardization.
- Conduct regular data audits, enforce strict access controls and maintain data interoperability across platforms.
- Establish strong data integrity SLAs with vendors to ensure consistent accuracy in all collaborative processes.

Protecting children's data and sensitive populations

Children mostly lack cognitive capability to understand long-term impact of their personal data. Certainly, children's data is critically important and requires special protection. Thereafter child-centric vigorous privacy framework must be in place and incorporated into all booking and passenger management systems.

- Enforce AI-Powered biometric systems to detect age anticipatory and ensure parental explicit consent, prior to children's personal data processing.
- Collect only essential information in addition to applying strong encryption and anonymization.
- Provide clear, family-friendly privacy notices.
- Conduct DPIAs specific to minors' data and completely align with international child privacy laws (such as COPPA and GDPR Art. 8).

Privacy-enhanced E-Governance and interagency coordination

In today's smart aviation environment, where emerging technologies are considered as critical factors for success, Governmental and airport digital systems must embed privacy-by-design and privacy-by-default in all services, in addition to the following:

- Mandate MFA (Multi-Factor Authentication), full-cycle enhanced encryption and regular vulnerability testing for all E-Governance platforms.
- Promote data sharing transparency among civil aviation, cybersecurity and border agencies.
- Deploy immutable auditing using blockchain or digital ledger systems for traceability of interagency data flows.
- Train staff regularly in ethical data handling and regulatory compliance.

Ethical and secure BPO

Outsourced operations increase data privacy risks if not tightly governed and strong privacy rules are not enforced. The subsequent initial measures must be established, across BPO framework.

- Conduct vendor due diligence with ISO 27001 certification checks and privacy audits, in addition to SOC-2 (Data Security and Operational Controls) report to increase trust.
- Enforce exhaustive DPAs, clearly defining ownership, retention and destruction policies.
- Apply data segmentation, encryption and enforce AI-Driven anomaly detection in all outsourced systems.
- Require vendors to fully align with GDPR standards, draft all-inclusive SCC - contracts that clearly define the accountabilities, responsibilities, privileges of all stakeholders and conduct joint risk assessments.
- Use tokenized identity data to prevent unauthorized access by third-party staff.

Privacy-preserving collaborative data processing

Aviation heavily relies on cross-entity collaboration (airlines, airports, regulators). To decrease the probability of data privacy risks, industry must adopt the following:

- Implement federated learning, homomorphic encryption and differential privacy for data sharing without exposure. He is a powerful cryptographic method that allows data to be processed and analyzed while still encrypted, without needing to decrypt it first.
- Establish unified accountability models and cooperative cybersecurity audits among stakeholders.
- Create private and restricted blockchain networks with controlled access for enhanced data traceability and control.
- Leverage AI-Enabled monitoring for anomaly and nonconformity detection across collaborative data ecosystems.

AI-driven privacy methods-a promising path forward

AI-driven privacy methods offer a strong and future ready solution to the growing challenges of big data landscape. These systems are highly scalable because they use automation, distributed architectures and smart resource management, allowing them to handle massive and complex datasets efficiently. They maintain data usefulness through adaptive anonymization, differential privacy and continuous optimization, meaning privacy can be protected without losing the real value of data.

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

In addition, AI models can automatically monitor compliance by detecting policy violations in real time, sending alerts and even adjusting privacy settings dynamically. Their context aware learning enables them to adapt as new threats appear, which is especially useful for constantly changing data environments like IoT, cloud storage and social media. By applying deep learning, anomaly detection and predictive modeling, these systems can identify hidden or emerging risks before they cause harm (Table 2).

Feature	Traditional privacy methods	AI-driven privacy methods
Scalability	Limited scalability in handling large or rapidly growing datasets due to manual processes and rigid infrastructure.	Highly scalable through automation, distributed architectures and intelligent resource optimization.
Data utility	Reduced data utility because of static anonymization and data suppression techniques that often degrade data quality.	Enhanced utility *via* adaptive anonymization, differential privacy and continuous optimization to preserve analytical value.
Compliance & monitoring	Manual, time-consuming compliance checks; often reactive to regulatory changes.	Automated compliance monitoring with real-time alerts, policy adaptation and continuous auditing based on AI-driven analytics.
Adaptability	Static, rule-based systems that struggle to adjust to new threats or evolving data environments.	Dynamic, context-aware algorithms that learn from patterns and adapt to emerging privacy risks.
Applications	Primarily effective in structured datasets and well-defined environments (e.g., relational databases).	Suitable for both structured and unstructured data across diverse sources (e.g., IoT, cloud, social media).
Processing speed	Slower processing due to human oversight and repetitive manual interventions.	Rapid processing and decision-making through Al automation, real-time detection and predictive analytics.
Transparency & explainability	High transparency but limited flexibility, users can audit simple rule sets easily.	Requires explainable Al frameworks to ensure interpretability and trust in algorithmic privacy decisions.
Threat detection	Relies on predefined rules; limited ability to detect novel or evolving privacy breaches.	Employs anomaly detection, predictive modeling and deep learning to identify new and complex privacy threats.
Cost efficiency	Higher operational costs due to manual labor and repetitive processes.	Long-term cost efficiency through automation, though initial deployment may require higher investment.
Examples	Data masking, pseudonymization, encryption and rule-based access controls.	Differential privacy, federated learning, homomorphic encryption and Al-driven risk scoring systems.

Table 2: Evolving paradigms - Traditional *vs.* AI-driven big data protection frameworks.

According to Williamson and Prybutok, “Integrating advanced technologies such as Homomorphic Encryption (HE), Differential Privacy (DP) and blockchain technology is innovative and essential in addressing the complex data security challenges” [14]. AI driven frameworks support all such advanced techniques in addition to federated learning, homomorphic encryption and AI-based risk scoring. Although its initial setup may involve higher costs, their automation and long-term efficiency, make them cost effective in the future. Importantly, the use of explainable AI ensures that privacy decisions remain transparent and trustworthy for users and regulators alike. In a nutshell, AI driven privacy protection combines intelligence, automation and adaptability to address modern data privacy needs in ways that traditional methods cannot [19-146].

Conclusion

Big Data Ecosystem has transformed industries' operations, enhanced productivity and overall performance. However, criticality is far from negligible because technology has also introduced high impact data privacy challenges at various stages of the big data lifecycle. This study highlighted the major components of big data privacy attacks and the dimensions of big data privacy i.e., legal, technical, political and ethical. The analysis showed that the data collection and sharing phases carry the highest probability of data privacy risks, specifically due to weak governance frameworks, inappropriate data protection controls, lack of employees’ awareness, unregulated data sharing and limited visibility in data exchange mechanisms.

The study’s primary focus was to identify foremost critical challenges pertaining to big data privacy in different industries. The analysis underlined the core challenges consist of data quality deficiencies, compromised data integrity, deviation from proper technological practices, non-compliant collaborative frameworks, nonconformance BPO, regulatory shortfall in E-Governance, compromised children’s data, unregulated automation and orchestration, imbalanced data profitability and data privacy, variation in regulatory frameworks, obscure and ambiguous consent, organizational data privacy failures and unethical data analysis practices. these challenges highlight key vulnerabilities and breaches that can lead organizations to several serious consequences ranging from legal, operational, reputational, brand damage, loss of trust, financial implications and strategic undesirable outcomes.

Addressing them is essential for strengthening overall data protection, privacy and ensuring compliance with national and global regulations. Implementing standardized data governance policies and privacy-by-design principles can significantly reduce these risks. To address these issues, the study proposes actionable guidelines that promote privacy-by-design, human-in-the-loop oversight, secure orchestration, real-time data integrity, immutable automated auditing, continuous monitoring, transparent and granular consent management, harmonizing regulatory disparities, privacy-enhanced E-Governance and interagency coordination, ethical and secure business process outsourcing, privacy-preserving collaborative data processing, regulatory alignment and cross-border data Governance. Implementing these guidelines can help organizations maintain trust, meet regulatory expectations and achieve sustainable digital growth.

Indeed, each of the afore-mentioned proposed techniques has trade-off (strength as well as weaknesses). Each proposed solution excels at solving certain challenges, but none is a one-size-fits-all

solution. Therefore, there is a pressing need for researchers to critically investigate each approach to determine most applicable path forward, since every proposed solution brings its own unique set of challenges. Considering the dynamic nature of big data environments, the following guidelines are provided for further exploration and comparative evaluation:

Data anonymization and re-identification

Even when personal information (like names, addresses, or mobiles) is removed from datasets, people can still be identified with the combination of distinct data sources. However, as of now, there are no tools or technologies that can completely measure and prevent this re-identification process, especially when multiple data sources are integrated and complex datasets are linked together. This creates a critical gap from data protection perspective because existing anonymization methods are not strong enough against modern data analysis techniques. There is a strong need to design new and advanced methods that can detect, test and at the very least, limit the risk of re-identification. Researchers should focus on developing tools that can analyze how easily data can be traced back to individuals and suggest ways to strengthen privacy protection across multiple datasets.

Data integrity and privacy leakage

API vulnerabilities, misconfiguration, external attacks and insider threat (malicious or negligent) can lead to critical data exposure. Even if data is encrypted, it can still leak private information through hidden channels such as timing information, data size, length, access patterns and frequency. Currently, there are no reliable systems to track where data originally comes from or to monitor how it is transformed, which makes it difficult to detect leaks. This lack of transparency in data flow causes major privacy and security challenges. There is a significant requirement for advanced tracking systems that can ensure both authenticity and privacy at the same time. Researchers should focus on creating frameworks that continuously monitor data movement, detect leaks automatically and verify that data remains private and unchanged during processing.

IoT and smart devices

Smart multifunction devices, such as security cameras, surveillance systems, biometric systems, sensors, meeting room systems and voice assistants, often share information across many interconnected systems. However, currently available privacy protections' solutions only work on individual devices and do not cover all interconnected devices collectively. This creates a gap in overall privacy because one weak device can compromise the security of all remaining. There is an essential demand to develop robust privacy framework, tools and technologies that can together protect information across all connected devices. These solutions should ensure that data remains safe even when devices communicate or share data within an IoT environment.

Wearable devices and health data

Wearables like fitness trackers and smartwatches collect highly personal and sensitive health data at high occurrence rates. At present, there are no clear technical or legal standards that define principally how this data should be stored, processed, shared, or deleted. This absence of standard guidelines leaves user information vulnerable to misuse. This high priority situation persuades both stricter legal policies and better technical measures to safeguard health data. New technologies and privacy rules must be created to ensure that users' information remains confidential, secure and under data subjects' control.

Cloud and cross border privacy

Cloud computing often stores user data in multiple countries and users rarely know where their information is stored. Many regulations require organizations to keep and process data within specific geographic or legal boundaries. There are very few tools that can verify whether companies follow privacy laws in different regions or not. This creates a trust gap between cloud users and service providers. There is an increasing requirement to build systems that link legal commitments to real technical enforcement. Such systems should deliver transparency, allowing users and regulators to verify that companies handle cross border data storage responsibly and according to privacy regulations.

Trusted and confidential computing

Although technologies generally claim to be secure but can still expose data through new forms of cyberattacks. There are no strong models to verify and prove the true level of privacy these systems propose. This makes it hard to trust them, specifically for sensitive data processing. Researchers are required to conduct deeper studies to make confidential computing more reliable and verifiable. The aim should be to develop testing frameworks that can evaluate privacy protection and ensure that these systems truly prevent sensitive information leaks.

Data privacy in blockchain

Technology keeps data permanently and is very useful from tracing and immutable auditing perspective. However, this immutability feature conflicts with data subjects' rights like the "right to be forgotten" and "right to rectification". Currently, there are very few practical ways to keep blockchain systems compatible with these legal rights, by some means. This gap limits blockchain's safe use in privacy sensitive fields. There is a big demand for continuous research to design systems that balance privacy rights with blockchain's permanent structure. Modern solutions must allow selective deletion, alteration and encryption of personal data without breaking blockchain's integrity.

Digital fingerprinting and tracking

Subsequent to traditional cookies being restricted, advanced hidden approaches to track online users, have already been placed. These include device fingerprinting and behavioral tracking, which are not yet well studied or understood. This lack of research leaves users unprotected from advanced forms of surveillance. It is extremely necessary to conduct detailed research on these modern tracking technologies to understand their long-term privacy impact. Regulations and new countermeasures must be developed to protect users from these invisible tracking methods.

Consent and user control

Most of the users swiftly agree to privacy terms without understanding what they mean. Current consent systems are long (multi-pages), complex and certainly ineffective in protecting user rights. This shows a major gap in how privacy consent is designed and managed. There is a crucial need to create easy to understand,

transparent and user-friendly consent mechanisms, which should help people make informed choices and maintain real control over how their personal data is collected and finally used.

Federated learning, differential privacy and ai training (LLM)

Large AI models like Gemini and ChatGPT are trained on massive datasets that can include private or personal information. Differential privacy helps protect individual data in shared datasets. At the same time in Federated Learning, AI systems learn from decentralized data stored on many devices. Together, federated learning and differential privacy form a privacy preserving framework, ideal for sensitive environments like aviation. While Differential Privacy mechanisms strengthen data privacy during AI model training, but adding too much DP noise makes model less accurate, whereas too little noise increases the probability of privacy risk. Moreover, there is no standard system to detect, remove, or stop leaks during AI learning process.

To address these serious risks, it is essential to build strong and reliable privacy safeguards (tools and technologies) and "machine unlearning" techniques that allow AI models to delete sensitive data. Research should focus on designing safer data handling methods for training large models. More practical testing and research are needed to create models that can maintain balance between privacy and data leverage. Researchers must find ways to secure communication channels and prevent any form of data leakage while keeping AI models accurate. The goal should also focus on designing tools that can apply differential privacy efficiently while keeping data meaningful for research and analysis.

Children's data protection

By way of example, online platforms often use face scans or ID uploads to verify age, which can risk children's privacy. There are no safe and trustworthy, privacy friendly tools for confirming age securely. This gap increases the chances of children's data being exposed or misused. There is an urgent need to research and develop anonymous, secure and nonintrusive age verification methods. These should protect children's safety without collecting unnecessary personal information.

Collaborative and third-party data processing

When multiple organizations share data, it is often unclear who is accountable and responsible for protecting it. There are no tools to continuously monitor whether other parties follow privacy rules or not. Even, there are no strong adaptable models that define data protection duties in collaborative environments. This creates confusion, a major gap in accountability and increases data privacy risks. Research is needed to build well-defined frameworks that establish trust, accountability as well as responsibility in data sharing partnerships. Secondly, automatic and continuous monitoring systems must be developed to ensure other parties handle data safely. Such systems should help ensure safety and fairness in collaborative data projects. These systems should also alert organizations to privacy violations and help maintain compliance.

Organizational privacy Governance

Many organizations have privacy programs, but they are not sure if these programs work or not. Although there are few mechanisms available like PPIs (Privacy Performance Indicators) and Privacy Maturing Models to assess how well privacy principles, controls and governance are functioning, none fit the requirement comprehensively. This makes it hard to compare or improve privacy practices. There is a critical need to research and create standard evaluation methods and common benchmarks that can evaluate systematically, how effective proposed privacy programs are. Indeed, this will help organizations strengthen their privacy governance, assurance about their privacy related projects and build overall user trust.

Global privacy law conflicts

Each country has its own privacy laws and they often conflict with one another. There is no unified global framework to bring these laws into harmony. This represents significant global challenge for companies that work internationally. Comprehensive in-depth research and cooperation are required to create robust international privacy standards that align with different national laws. Such coordination will support global data protection and simplify compliance for businesses.

Explainable AI and privacy

XAI (Explainability) enhances regulatory compliance by making AI decisions transparent, interpretable and auditable, as organizations can document AI behavior for regulators, internal audits, or compliance checks. Additionally, XAI also supports fairness (detects and mitigates bias), accountability (explains decisions for audits and responsibility) and consent management (informs users how data is used). Simultaneously, generating explanations often requires accessing, processing, or exposing sensitive personal information, which can inadvertently cause significant risks of unintended disclosure, misuse, or re-identification. Even as interpretability mechanisms enhance trust and governance, they must be carefully balanced with robust data protection measures. Researchers must develop techniques that allow clear explanations of AI behavior without risking sensitive data exposure.

Conflict of interest

The author declares no conflict of interest.

References

1. Demirol D, Das R, Hanbay D (2023) A key review on security and privacy of big data: Issues, challenges and future research directions. Signal, Image and Video Processing 17(4): 1335-1343. [Crossref] [GoogleScholar]
2. Asaad RR, Abdulnabi LN (2022) A Review on big data analytics between security and privacy issue. Academic Journal of Nawroz University 11(3): 178-184. [Crossref] [GoogleScholar]
3. Momin S, Avghade S, Chavan S (2023) Data security and privacy protection in web in Indian environment. International Journal of Advanced Research in Science, Communication and Technology 3(1): 144-150. [Crossref] [GoogleScholar]
4. Zhou Z, Wei Z, Wang F (2023) Privacy protection and data security for big data encryption: A review. In Proceedings of the 2023 3rd International Conference on Big Data, Artificial Intelligence and Risk Management 1087-1091. [Crossref] [Google Scholar]
5. Kajunju E, Hamilton M (2021) Health care privacy on the ground: Key challenges and practical solutions.
6. Ijaiya H (2024) Harnessing AI for data privacy: Examining risks, opportunities and strategic future directions. International Journal of

Science and Research Archive 13(2): 2878-2892. [Crossref] [GoogleScholar]

7. Voss WG (2021) Airline commercial use of EU personal data in the context of the GDPR. British Airways and Schrems II 19(2): 377. [Google Scholar]
8. Ijaiya H (2024) Balancing data privacy and technology advancements: Navigating ethical challenges and shaping policy solutions. In International Journal of Research Publication and Reviews 5(11): 8118-8130. [GoogleScholar]
9. Mulder T, Vellinga NE (2021) Exploring data protection challenges of automated driving. Computer Law and Security Review 40. [Crossref] [GoogleScholar]
10. Oluwatosin R, Eneh NE, Ehimuan B, Anyanwu A, Olorunsogo T, et al. (2024) Privacy law challenges in the digital age: A global review of legislation and enforcement. International Journal of Applied Research in Social Sciences 6(1): 73-88. [Crossref]
11. Oluwabunmi L, Naiho HNN, Adeleke GS, Udeh EO, Labake TT (2024) Data privacy and security challenges in environmental research: Approaches to safeguarding sensitive information. International Journal of Applied Research in Social Sciences, 6(6): 1193–1214. [Crossref] [GoogleScholar]
12. Goel P, Patel R, Garg D (2021). A review on big data: Privacy and security challenges. IEEE 705-709. [Crossref] [GoogleScholar]
13. Shahriar S, Dara R, Akalu R (2025) A comprehensive review of current trends, challenges and opportunities in text data privacy. Computers and Security, 151. [Crossref] [GoogleScholar]
14. Williamson SM, Prybutok V (2024) Balancing privacy and progress: A Review of privacy challenges, systemic oversight and patient perceptions in AI-driven healthcare. Applied Sciences 14(2): 675. [Crossref] [Google Scholar]
15. Hasanzadeh S, Zhou W (2025) Wearable technologies in construction: Age-related perceptions, privacy concerns and feedback preferences. CIB Conferences 1(1): 116. [Crossref] [Google Scholar]
16. Ramya S, Devi SR, Pandian SP, Suguna G, Suganya R, et al. (2023) Analyzing big data challenges and security issues in data privacy. International Research Journal of Modernization in Engineering 421.
17. Taherdoost H (2023) Navigating the ethical and privacy concerns of big data and machine learning in decision making. Intelligent and Converged Networks 4(4): 280-295. [Crossref] [Google Scholar]
18. McCausland R (2024) White paper: Data protection and international carriage by air.
19. Ahmed A (2025) Google processes 158,548 searches every second, here's how much it adds up to in just one year. Digital Information World.
20. Mohamed HA, Al-Azab MR (2021) Big data analytics in airlines: Opportunities and challenges. Journal of Association of Arab Universities for Tourism and Hospitality 21(4): 77-112. [Crossref] [GoogleScolar]
21. Ahmadi MSSA, Alhejaili FM, Alamry MS, (2024) Comprehensive review of big data analytics, health information systems and data privacy ethics. Journal of Ecohumanism 3(8): 5510-5518. [Crossref]
22. Ahmed H, Khan H, Khan MA (2023) A survey on security and privacy Of Automatic Dependent Surveillance—Broadcast (ADS-B) protocol: Challenges, potential solutions and future directions. TechRxiv. [Crossref] [GoogleScholar]
23. Alabdulatif A, Thilakarathne NN, Kalinaki K (2023) A novel cloud enabled access control model for preserving the security and privacy of medical big data. Electronics 12(12): 2646. [Crossref] [GoogleScholar]
24. Ali TA, Khafagy MH, Farrag MH (2022) Special Negative Database (SNDB) for protecting privacy in big data. International Journal of Advanced Computer Science and Applications 13(1): 1-14. [Crossref] [GoogleScholar]
25. Al-Zahrani A, Al-Hebbi M (2022) Big data major security issues: Challenges and defense strategies. Tehnicki Glasnik 16(2): 197-204. [Crossref] [GoogleScholar]
26. Amiri-Zarandi M, Dara RA, Duncan E, Fraser ED (2022) Big data privacy in smart farming: A review. Sustainability 14(15): 9120. [Crossref] [GoogleScholar]
27. Anjum N, Latif Z, Chen H (2025) Security and privacy of industrial big data: Motivation, opportunities and challenges. Journal of Network and Computer Applications 237: 104130. [Crossref] [GoogleScholar]
28. Nirmalgowri K, Vidhya A (2025) Blockchain and big data: A hybrid model for ensuring security and privacy in IOT and social media– a review. Singaporean Journal of Scientific Research (SJSR) An International Journal (AIJ) 17(1): 41-52. [GoogleScholar]
29. Badawy M, Ramadan N, Hefny HA (2024) Big Data analytics in healthcare: Data sources, tools, challenges and opportunities. Journal of Electrical Systems and Information Technology 11(1): 63. [Crossref] [GoogleScholar]
30. Baig HA (2022) A protection layer over MapReduce framework for big data privacy. International Journal of Computer and Information Technology 11(2). [Crossref]
31. Belani S, Tiarks GC, Mookerjee N, Rajput V (2021) " I agree to disagree": Comparative ethical and legal analysis of big data and genomics for privacy, consent and ownership. Cureus 13(10): e18736. [Crossref] [GoogleScholar] [Pubmed]
32. Ehimuan B, Chimezie OOb, Akagha OV, Reis O, Oguejiofor BB (2024) Global data privacy laws: A critical review of technology's impact on user rights. World Journal of Advanced Research and Reviews, 21(2): 1058-1070. [Crossref]
33. Biswas S, Fole A, Khare N, Agrawal P (2023) Enhancing correlated Big Data privacy using differential privacy and machine learning. Journal of Big Data, 10(1): 30. [Crossref] [GoogleScholar]
34. Biswas S, Khare N, Agrawal P, Jain P (2021) Machine learning concepts for correlated big data privacy. Journal of Big Data 8(1): 157. [Crossref] [GoogleScholar]
35. Blumenstock JE, Kohli N (2023) Big Data privacy in emerging market fintech and financial services: A research agenda. arXiv Preprint. [Crossref] [GoogleScholar]
36. Chakkappan G, Morshed A (2024) Explainable AI and big data analytics for data security risk and privacy issues in the financial industry. 2024 IEEE Conference. [Crossref] [GoogleScholar]
37. Chandrakala P, Balmiki V, Upadhyay D (2023) Privacy-preserving techniques in big data analytics: A cybersecurity assessment. IEEE 282-286. [Crossref] [GoogleScholar]
38. Chandrakar I, Hulipalled VR (2022) Improved technique for preserving privacy while mining real time big data. International Journal of Communication Networks and Information Security 14(1): 86-92. [GoogleScholar]
39. Chang V, Ji Z, Arami M () Privacy and ethical issues of big data in the airline industry. SciTePress 139-148. [GoogleScholar]
40. Chen H, Babar MA (2024) Security for machine learning-based software systems: A survey of threats, practices and challenges. ACM Computing Surveys 56(6): 1-38. [Crossref] [GoogleScholar]
41. Choo KKR, Dehghantanha A (2020) Handbook of big data privacy. In Handbook of Big Data Privacy. Springer International Publishing. [Crossref] [GoogleScholar]
42. Cuzzocrea A (2025) Big data privacy in multidimensional domains: Models, issues, paradigms. 2025 IEEE 49th Annual Computers, Software. [Crossref] [GoogleScholar]

43. Cuzzocrea A, Soufargi S (2025) Privacy-preserving multidimensional big data analytics models, methods and techniques: A comprehensive survey. Expert Systems with Applications 270: 126387. [Crossref] [GoogleScholar]
44. Debad SJ, Ganz A, Snyder M (2024) Wearing your heart (monitor) on your sleeve: Will data be the new doctor? Frontiers for Young Minds 12. [Crossref] [GoogleScholar]
45. Deepika R, Valarmathi K (2022) Privacy and security of big data-a high perspective investigation. IEEE. [Crossref] [GoogleScholar]
46. Djumadin Z (2023) Privacy protection in the big data era: A review of personal data protection policies. Restoration Journal: Law and Politics 1(2): 72-78. [GoogleScholar]
47. Ebert I, Wildhaber I, Adams-Prassl J (2021) Big data in the workplace: Privacy due diligence as a human rights-based approach to employee privacy protection. Big Data & Society 8(1): 20539517211013051. [Crossref] [GoogleScholar]
48. Adaga EM, Okorie GN, Egieya ZE, Ikwue U, Udeh CA, et al. (2024) The role of big data in business strategy: A critical review. Computer Science & IT Research Journal 4(3): 327-350. [Crossref] [GoogleScholar]
49. Eling M, Gemmo I, Guxha D, Schmeiser H (2024) Big data, risk classification, and privacy in insurance markets. The Geneva Risk and Insurance Review 49(1): 75-126. [Crossref] [GoogleScholar]
50. Elkawkagy M, Elwan E, Alsumayt A, Elbeh H, Aljameel S (2024) Elevating big data privacy: Innovative strategies and challenges in data abundance. IEEE 12: 20931-20941. [Crossref] [GoogleScholar]
51. Asituha E (2024) A comprehensive overview of privacy, security and performance issues in flying Ad Hoc networks. World Journal of Advanced Research and Reviews 23(1): 1902-1930. [Crossref]
52. Fashakh A, Abddulkader H (2022) Big data and cybersecurity: A review of key privacy and security challenges. IEEE 1-7. [Crossref] [GoogleScholar]
53. Ferradi H, Cao J, Jiang S, Cao Y, Saxena D (2022) Security and privacy in big data sharing: State-of-the-art and research directions. arXiv Preprint arXiv. [Crossref] [GoogleScholar]
54. Filaly Y, Berros N, Bouzekri YEEL (2025) A comprehensive survey on big data privacy and Hadoop security: Insights into encryption mechanisms and emerging trends. Results in Engineering 27: 106203. [Crossref] [GoogleScholar]
55. Fortino G, Liotta A (2026) Internet of things technology, communications and computing series editors.
56. Funde S, Swain G (2022) Big data privacy and security using abundant data recovery techniques and data obliviousness methodologies. IEEE 10: 105458-105484. [Crossref] [GoogleScholar]
57. Georgiadis G, Poels G (2022) Towards a privacy impact assessment methodology to support the requirements of the general data protection regulation in a big data analytics context: A systematic literature review. Computer Law &Security Review 44: 105640. [Crossref] [GoogleScholar]
58. Okorie GN, Udeh CA, Adaga EM, DaraOjimba OD, Oriekhoe OI (2024) Ethical considerations in data collection and analysis: A review: Investigating ethical practices and challenges in modern data collection and analysis. International Journal of Applied Research in Social Sciences 6(1). [Crossref]
59. Greenleaf G (2023) Global tables of data privacy laws and bills. 8th edn. SSRN. [GoogleScholar]
60. Hadi HJ, Cao Y, Nisa KU, Jamil AM, Ni Q (2023) A comprehensive survey on security, privacy issues and emerging defence technologies for UAVs. Journal of Network and Computer Applications 213: 103607. [Crossref] [GoogleScholar]
61. Hamad F, Fakhuri H, Jabbar AS (2022) Big data opportunities and challenges for analytics strategies in Jordanian academic libraries. New Review of Academic Librarianship 28(1): 37-60. [Crossref] [GoogleScholar]
62. Harper W (2022) Big data security and privacy in cloud computing. ResearchGate.
63. He M, Chen Y (2024) Personal data protection in China: Progress, challenges and prospects in the age of big data and AI. Telecommunications Policy 49(10): 103076. [Crossref] [GoogleScholar]
64. Herriger C, Merlo O, Eisingerich AB, Arigayota AR (2025) Context-contingent privacy concerns and exploration of the privacy paradox in the age of AI, augmented reality, big data and the internet of things. Journal of Medical Internet 27: e71951. [Crossref] [GoogleScholar]
65. Hye-Yoon P, Yoon EJ (2025) An study on ethical collection and use of airline tourism data. Journal of Research and Publication Ethics 6(1): 13-19. [Crossref] [GoogleScholar]
66. Idoko IP, Igbede MA, Manuel HNN (2024) Big data and AI in employment: The dual challenge of workforce replacement and protecting customer privacy in biometric data usage. Global Journal of Engineering and Technology Advances. [Crossref] [GoogleScholar]
67. James M (2024) The ethical and legal implications of using big data and artificial intelligence for public relations campaigns in the United States. International Journal of Communication and Public Relation 9(1): 38-52. [Crossref]
68. Ngesa J (2023) Tackling security and privacy challenges in the realm of big data analytics. World Journal of Advanced Research and Reviews 21(2) 552–576. [Crossref] [GoogleScholar]
69. Jenkins T (2024) Classification algorithms for high-dimensional big data with privacy constraints. researchgate.net. [GoogleScholar]
70. Arowoogun JO, Babawarun O, Chidi R, Adeniyi AO, Okolo CA (2024) A comprehensive review of data analytics in healthcare management: Leveraging big data for decision-making. World Journal of Advanced Research and Reviews 21(2): 1810-1821. [Crossref] [GoogleScholar]
71. Joshi D, Sanghi A, Agarwal G (2024) Techniques for protecting privacy in big data security: A comprehensive review. Conference on Electrical. [Crossref] [GoogleScholar]
72. Journal I (2024) Big data analytics: To understand the challenges. International Journal of Scientific Research In Engineering and Management 8(1).
73. Kabashkin I (2025) Framework for addressing imbalanced data in aviation with federated learning. Information (Switzerland) 16(2). [Crossref] [GoogleScholar]
74. Kartheek P (2024) Security and privacy technique in big data: A review. philpapers.org.
75. Kasera S, Gehlot A, Uniyal V, Pandey S (2023) Right to digital privacy: A technological intervention of blockchain and big data analytics. On Innovative Data. [Crossref] [GoogleScholar]
76. Katamoura S, Aksoy MS (2024) Privacy and security in artificial intelligence and machine learning systems for renewable energy Big Data. 2024 21st Learning and. [Crossref] [GoogleScholar]
77. Kemp S (2025) Digital 2025 July Global stats hot report. DataReportal. Retrieved from.
78. Khalid A (2023) The role of differential privacy in big data: Applications and challenges. Journal of Big Data Privacy Management. [GoogleScholar]
79. Khalil MK, Ahmed A, Al Amri S, Khalid M, Latif A, et al. (n.d.-a) Data Privacy and Security Challenges in Big Data Analytics: A Review of Current Solutions and Future Directions.
80. Khan J, Ahmad N (2023) Security and privacy technique in big data: A review. 2023 10th International Conference on.
81. Khan MA, Gupta P, Sultan AA, Singh P, Shivam, et al. (2024) Security in cloud computing: Issues and challenges. International Journal of Intelligent Systems and Applications in Engineering 12(17): 674-681.
82. Khurana J (2025) Security privacy challenges of big data over cloud: A review. Progressive Computational Intelligence, Information. [GoogleScholar]

83. Kumar R (2022) A Review on big data: Privacy and security challenges. AG Volumes. [GoogleScholar]
84. Li K, Wang Z, Wang Y, Luo B, Li F (2021) Poster: Ethics of computer security and privacy research—current status and trends from a data perspective. CCS 2023 - Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security 3558-3560. [Crossref] [GoogleScholar]
85. Liang J, Dissertation A (2023) Rsk interpretation of differential privacy and its applications. [GoogleScholar]
86. Liu J (2023) An overview of big data mining and data privacy protection technologies. Applied and Computational Engineering 21: 187-192. [Crossref]
87. Lopes NM, Aparicio M, Neves FT (2025) Challenges and prospects of artificial intelligence in aviation: A bibliometric study. Data Science and Management 8(2): 207-223. [Crossref] [GoogleScholar]
88. C-Shan L, Leung YT, Yang, et al. (2022) Proceedings of the International Forum on Shipping, Ports and Airports (IFSPA) 2022: Resilience from adversity. Hong Kong. C.Y. Tung International Centre for Maritime Studies, Department of Logistics and Maritime Studies, the Hong Kong Polytechnic University.
89. Madavarapu JB, Nainwal A, Shnain AH (2024) Federated learning for privacy-preserving medical data analytics in big data. IEEE. [Crossref] [GoogleScholar]
90. Manroop L, Malik A, Milner M (2024) The ethical implications of big data in human resource management. Human Resource Management Review 34(2): 101012. [Crossref] [GoogleScholar]
91. Martin C (2024) Algorithms for privacy-aware data mining in big data systems. researchgate.net. [GoogleScholar]
92. Pisal T, Jadhav V (2024) The use of big data to improve disaster response and preparedness efforts. International Journal of Advanced Research in Science, Communication and Technology.
93. NPS, D Usha (2023) A review on big data privacy and security in health care. International Journal of Engineering Research and Sustainable Technologies (IJERST) 1(1): 38-49. [Crossref] [GoogleScholar]
94. Nageshwaran G, Harris RC (2021) Review of the role of big data and digital technologies in controlling COVID-19 in Asia: Public health interest *vs*. Privacy. Digital. [Crossref] [GoogleScholar]
95. Naretto F, Monreale A, Giannotti F (2025) Evaluating the privacy exposure of interpretable Global and local explainers. In TRANSACTIONS ON DATA PRIVACY 18: 67-93. [GoogleScholar]
96. Nguyen TM, Xuan-Son VU (2023) Privacy and trust in IoT ecosystems with big data: A survey of perspectives and challenges. International Conference on Big Data. [Crossref] [GoogleScholar]
97. NP, HK, Prabhudeva S (2022) An authorization framework for preserving privacy of big medical data *via* blockchain in cloud server. International Journal of Advanced 13(3): 1-12. [Crossref] [GoogleScholar]
98. Nurrizqa and Nurrisma (2024) Big data analytics for decision support in healthcare information systems opportunities and challenges. Journal Informatic, Education and Management (JIEM) 6(1). [Crossref]
99. OBIANYO, CI, Ater SV (2023) The legal implication of data privacy and protection in the age of big data and the ever-emerging technologies in Nigeria. Idemili Bar Journal. [GoogleScholar]
100. Ogbuke NJ, Yusuf YY, Dharma K (2022) Big data supply chain analytics: Ethical, privacy and security challenges posed to business, industries and society. Production Planning and control. [Crossref] [GoogleScholar]
101. Ajala OA, Arinze CA, Ofodile OC, Okoye CC, Daraojimba OD (2024) Reviewing advancements in privacy-enhancing technologies for big data analytics in an era of increased surveillance. World Journal of Advanced Engineering Technology and Sciences 11(1). [Crossref] [GoogleScholar]
102. Joel OS, Oyewole AT, Odunaiya OG, Soyombo OT (2024) The impact of digital transformation on business development strategies: Trends, challenges and opportunities analyzed. World Journal of Advanced Research and Reviews 21(3): 617-624. [Crossref] [GoogleScholar]
103. Odeyemi O, Mhlongo NZ, Nwankwo EE, Scholatica UC, Okoye CC (2024) Big Data applications in portfolio management: A review of techniques and strategies. International Journal of Science and Research Archive 11(1): 1984-1999. [Crossref] [GoogleScholar]
104. Oracle (2021) What is big data?
105. Pamarthi K (2024) Security and privacy technique in big data: A review. North American Journal of Engineering Research 5(1).
106. Pramanik MI, Lau RYK, Hossain MS (2021) Privacy preserving big data analytics: A critical analysis of state-of-the-art. Data Mining and Knowledge Discovery [Crossref] [GoogleScholar]
107. Puri GD, Haritha D (2023) Implementation of big data privacy preservation technique for electronic health records in multivendor environment. International Journal of Advanced Computer Science and Applications 14(2). [GoogleScholar]
108. Qiang Q, Wang H, Xie C (2023) Big data and privacy: A review on the effectiveness of current data privacy protection strategies. fe-vision.com. [GoogleScholar]
109. Rafiq F, Awan MJ, Yasin A, Nobanee H (2022) Privacy prevention of big data applications: A systematic literature review. SAGE 12(2). [Crossref] [GoogleScholar]
110. Raheem B (2021) Big Data Security and Privacy in Cloud Computing. Researchgate.net.
111. Rajan, AA, VV (2024) Systematic survey: Secure and privacy-preserving big data analytics in cloud. Journal of Computer Information Systems. [Crossref] [GoogleScholar]
112. Ramakrishnan R, Sujithra R, Niranjalin AJ (2024) Privacy challenges and solutions in big data analytics: A comprehensive review. International Journal for Research in Applied Science and Engineering Technology 12(5): 2435-2440. [Crossref]
113. Raza A (2023) Advances in secure data sharing for big data privacy preservation. Journal of Big Data Privacy Management. [GoogleScholar]
114. Raza A (2025) Predicting privacy risks in big data environments: A machine learning approach. Journal of Big Data Privacy Management. [GoogleScholar]
115. Raza MA (2023) Privacy management in financial data: Strategies for securing big data in the banking sector. Journal of Big Data Privacy Management. [GoogleScholar]
116. Reka SS, Dragicevic T, Venugopal P, Ravi V (2024) Big data analytics and artificial intelligence aspects for privacy and security concerns for demand response modelling in smart grid: A futuristic approach. Heliyon. [GoogleScholar]
117. Rubinstein IS (2013) Big data: The end of privacy or a new beginning?. International Data Privacy Law 3:74. [Crossref] [GoogleScholar]
118. Saidah M, Maylaffayza H (2024) Data privacy protection in Islamic communication perspective. Komunika: Jurnal Dakwah Dan Komunikasi 18(1). [Crossref]
119. Saura JR, Ribeiro-Soriano D, Palacios-Marqués D (2021) Setting privacy “by default” in social IoT: Theorizing the challenges and directions in Big Data Research. Big Data Research. [Crossref] [GoogleScholar]
120. Shahzad H, Veliky C, Le H, Qureshi S, Phillips FM (2024) Preserving privacy in big data research: The role of federated learning in spine surgery. European Spine. [Crossref] [GoogleScholar]
121. Shanthi R, Babu MD, Kousika N, Vijayaraj C, Choubey SB, et al. (2024) Advanced privacy-preserving framework using homomorphic encryption and adaptive privacy parameters for scalable big data analysis. International Journal of Intelligent Systems and Applications in Engineering 12(11): 160-165.
122. Sharma I, Thakur R (2023) Critical data encryption: A deep study.
123. Sharma K, Baalamurugan KM (2021) A review on big data privacy and security techniques for the healthcare records. 2021 3rd International. [Crossref] [GoogleScholar]
124. Shin H, Ryu K, Kim JY, Lee S (2024) Application of privacy protection technology to healthcare big data. Digital Health. [Crossref] [GoogleScholar]
125. Sholihati ID, Wedha BY, Ningsih S, Sari RTK (2024) The Impact of big data on enterprise architectural design: A conceptual review. Journal of

Computer Networks, Architecture and High-Performance Computing 6(1). [Crossref] [Google Scholar]
126. Siddiqui A (2024) Privacy threats in big data: Identifying and mitigating vulnerabilities in data-driven systems. Journal of Big Data Privacy Management 2(1): 42-49. [Google Scholar]
127. Smirnova Y, Travieso-Morales V (2024) Understanding challenges of GDPR implementation in business enterprises: A systematic literature review. International Journal of Law and Management 66(3): 326-344. [Crossref] [Google Scholar]
128. Song B, Deng M, Pokhrel SR, Lan Q, Doss R, et al. (2025) Digital privacy under attack: Challenges and enablers. ACM Computing Surveys 58(4): 1-35. [Crossref] [Google Scholar]
129. Su P (2024) Privacy protection in mobile big data: Challenges and solutions. International Journal of Interactive Mobile Technologies 18(18): 49-61. [Crossref] [Google Scholar]
130. Thasna CA, Chawla M, Tiwari N (2023) A review of traditional and neural network methods for protecting privacy in big data analytics. International Conference on Deep Learning, Artificial Intelligence and Robotics 1001: 158-167. [Google Scholar]
131. Tian H, Presa-Reyes M, Tao Y, Wang T, Pouyanfar S, et al. (2022) Data analytics for air travel data: A survey and new perspectives. ACM Computing Surveys 54(8): 1-35. [Crossref] [Google Scholar]
132. Tripathi K, Biswas S, Khare N, Shukla S (2024) Tackling privacy concerns in correlated big data: A comprehensive review with machine learning insights. 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science: 1-6. [Crossref] [Google Scholar]
133. Trust Arc. The Ultimate Guide to Understanding Online Tracker Technology.
134. Usha D (2023) A review on big data privacy and security in health care. International Journal of Engineering Research and Sustainable Technologies 1(1): 38-49. [Crossref] [Google Scholar]
135. Vasa J, Thakkar A (2023) Deep learning: Differential privacy preservation in the era of big data. Journal of Computer Information Systems 63(3): 608-631. [Crossref] [Google Scholar]
136. Venish L (2023) Privacy-preserving techniques in big data analytics: Challenges and opportunities. researchgate.net.
137. Wadhwa B, Tomar P (2022) Security and privacy challenges in big data. Proceedings of Third Doctoral Symposium on Computational Intelligence 479: 821-837. [Google Scholar]
138. Wan H (2024) Overview of big data information privacy protection technology and challenges. Science and Technology of Engineering, Chemistry and Environmental Protection 1(5). [Crossref]
139. Wang H, Liu T (2024) Application of big data in tourism destination management: A case study of Changsha city. Urban Studies and Public Administration 7(1). [Crossref]
140. Wang S, Wei Z, Wang F (2024) Data security and privacy protection in the era of big data: A review. 2024 2nd International Conference on Computer Network Technology and Electronic and Information Engineering :73-77. [Crossref] [Google Scholar]
141. Wang Y (2024) realistic dilemma and path optimization of online personal information protection. Lecture Notes in Education Psychology and Public Media 39(1): 187-192. [Crossref]
142. Yuan ZS, Singh MM (2025) Privacy risks in health big data: A systematic literature review. [Google Scholar]
143. Zaabi MA, Alhashmi SM (2024) Big data security and privacy in healthcare: A systematic review and future research directions. Information Development [Crossref] [Google Scholar]
144. Zhang X, Yao Y, Wang Y (2024) Lightweight privacy protection scheme for cloud audit. *Journal of Chongqing University* 47(2): 75-83. [Google Scholar]
145. Zhang Y, Sun M (2023) Research progress on data security and privacy protection of wearable devices in the context of healthcare big data. Proceedings of the 2023 4th International Symposium on Artificial Intelligence for Medicine Science 965-968. [Crossref] [Google Scholar]
146. Zhao F (2024) Big data applications and mining in the healthcare field. Journal of Computing and Electronic Information Management 12(1). [Google Scholar]

Supplementary Information

Aspect / dimension	Present (big data era)	Past (pre-big data era)
Data Volume & Speed	Massive Volumes Near-Real Time Streaming Petabytes → Zettabytes	Smaller Volumes, Batch Processing; Slower Updates
Data Variety & Structure	Huge Semi-Structured, or Unstructured Data IoTs, Logs, social media, Videos, Images, Sensors	Mainly Structured Data RDBMS
Data Sources / Devices	Multitude: Smartphones, Wearables, Sensors, Webcams, Smart-Home Devices, social media, Cloud Services	Limited: Desktops, Laptops, Phone Calls, Paper Forms
User Awareness & Control	Users Less Aware Implicit Tracking, Consent Often “Bundled” or Vague	Users more aware of when & what was being collected
Transparency of Use	Less Transparent AI/ML, Complex Data Flows Pipelines and Integrations	Generally, More Transparent Organizations Disclosed Purpose of Data Collection
Regulatory Environment	Regulatory Disparities Cross Board Data Transfer Numerous National and International Data Protection Laws PDPL, GDPR, CCPA	Fewer Laws Mostly Local or National; Limited Cross Border Concerns
Risk of Re-identification & Inference	Higher Risk: Ability to Correlate Diverse Datasets Infer Identities or Sensitive Traits even from Anonymous Data	Lower Risk Easier Anonymization Fewer Linkage Points
Security Threats & Attack Surface	Much larger Attack Surface Sophisticated Cyber Attacks Insider Threats, Cloud/Hybrid Environments Introduce Further Vulnerabilities	Smaller Attack Surface Simpler Threats

https://doi.org/10.70844/ijas.2025.2.42

Innovative Journal of Applied Science

Impact of Breaches	Massive Scale Millions/Billions of Records Potential for Cascading Effects e.g. Identity Theft, Financial Loss, Reputation Damage	Localized Damage Usually, Fewer Records Compromised
Data Ownership & Stewardship	Blurred Ownership Users' Data Shared, Sold or Reused Third Parties, Brokers, AI Systems Involvement	Clearer Lines Data Collected by Organizations, Less Sharing
Ethical & Social Implications	Major Concern Bias, Discrimination, Surveillance, Pressure on Autonomy Issues of Fairness, Justice, Misuse by Governments or Corporations	Emerging, Less Discussed
Privacy related Tools	Advanced & Emerging Tools Privacy by Design Differential Privacy Federated Learning Zero-Knowledge-Proofs	Basic Tools Encryption, Access control, Pseudonymization
Governance & Accountability	Complex Governance Regulatory Oversights Internal and External Audits Public Scrutiny	Simple Governance Limited Audit, and Oversight
Compliance & Penalties	Strict Compliance Heavy Fines and Penalties Reputational Cost	Compliance Less Strict Light or Rare Penalties

Table S1: Big data impact - comparing then and now.

Disclaimer

The information contained in this article is provided for general informational and educational purposes only. While every effort has been made to ensure the accuracy, reliability, and completeness of the information presented, the author makes no representations or warranties, express or implied, regarding the accuracy, applicability, fitness, or completeness of the content. Readers are encouraged to verify facts independently and consult qualified professionals before making any decisions based on the information provided herein.

The opinions and interpretations expressed are those of the author alone and do not represent the official stance of any institution, organization, or entity with which the author may be affiliated. This article does not constitute legal, regulatory, financial, or compliance advice, and no attorney–client, advisor–client, or fiduciary relationship is created by reading or relying on this content. For specific legal or compliance-related questions, readers should seek professional counsel from qualified experts in data protection law or relevant regulatory authorities.

All examples, case studies, and scenarios mentioned are illustrative and may not reflect current legal frameworks, technological standards, or organizational practices. References to laws such as the General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA), or any other regional privacy legislation are made for explanatory purposes only and do not substitute official legal texts or legal interpretations.

The author expressly disclaims any responsibility or liability for any direct, indirect, incidental, consequential, or special damages arising out of or in any way connected with the use of, reliance upon, or performance of the information contained in this article. Any actions taken based on the content are undertaken at the reader's own discretion and risk.

Where this article includes links, citations, or references to external websites, research papers, or Third-Party resources, such inclusion does not imply endorsement or responsibility for the accuracy, reliability, or legality of external content. The author does not control and cannot guarantee the ongoing availability, accuracy, or integrity of external sources.

By accessing or using this article, readers acknowledge and agree that all information is provided "as is" and that the author bears no responsibility for any interpretation, decision, or action made in reliance upon the material herein. The author reserves the right to modify, update, or withdraw the content at any time without prior notice to ensure ongoing relevance and compliance with evolving privacy standards and regulations.