Decentralization and Federated Approach for Personal Data Protection and Privacy Control

Suyogita Singh[1*], Satya Bhushan Verma[2], Bineet Kumar Gupta[3], Anil Kumar Pandey[4], Surya Bhushan Dubey[5]

[1*,.2,.3,.4]Shri Ramswaroop Memorial University, Barabanki, Uttar Pradesh, India, 225003
[1]suyogitasingh0885@gmail.com,Satyabverma1@gmail.com,[3]bkguptacs@gmail.com,[4]anipandey@gmail.com
[5]Electrical Engineering, SR Group Of Institutions Lucknow, [5]suryadubey1@gmail.com

Abstract:
As the digital ecosystem evolves, personal data protection and privacy control have become paramount concerns. Traditional centralized approaches to data management and privacy often pose risks to individuals' data security and privacy. This research paper explores the potential of decentralization and federated approaches as innovative solutions for enhancing personal data protection and privacy control. By distributing data across multiple entities and empowering individuals with control over their data, decentralized and federated models offer increased security, privacy, and data autonomy. This paper investigates the benefits, challenges, and future prospects of decentralized and federated approaches in personal data protection and privacy control.

1 Introduction

The volume of private data being collected is experiencing a significant surge. Businesses and authorities leverage these data profiles to understand individuals, shape their opinions, and influence their behaviour. This can lead to more personalized experiences, tailored offerings, or more efficient resource utilization. However, it also introduces the risk of deception and misuse by the entities acquiring the data and individuals attempting to exploit it. In response to the rise in cybercrime and increasing consumer concerns over the control of personal information, efforts to create and implement user privacy measures are underway. For businesses dealing with personal information, the expenses related to data storage and security are on the rise. Living in a data-driven world has benefits for everyone, but the growing public concern revolves around user privacy. Individuals face an escalating risk of data theft or exploitation, which could have legal, financial, and reputational consequences. Centralized public and corporate entities accumulate vast amounts of private and sensitive information, leaving individuals with little to no control over the storage and usage of their personal data. Recent privacy-related incidents have gained widespread media attention, such as government surveillance and Facebook's extensive scientific study conducted without explicit participant notification [1-3]. In 2008, Nakamoto introduced bitcoin through a white paper, introducing blockchain, a decentralized public database that powers the virtual currency Bitcoin. The current status of blockchain is sometimes compared to the early days of the Internet in the mid-1990s, when its value and potential were not fully understood. However, as the importance of blockchain technology has become apparent, some countries have established research facilities in this field. Numerous studies have been conducted to comprehend the vast potential of blockchain in various smart settings, including business, healthcare, smart cities, and transportation. Blockchain technology has diverse applications, and numerous papers have been dedicated to exploring these uses. Federated Learning (FL) emerges as a well-known data analysis method for safety-critical systems like smart healthcare, industrial settings, and smart cities. FL, a prominent artificial intelligence strategy, has gained attention in recent years due to its specialized and advanced machine learning methodology. FL has the capability to retain all datasets locally, renew and update essential training settings, and split selected data records across actual or virtual machines in each federated zone. This approach enables datasets to uncover unstructured patterns using high-speed procedures and secure settings, ensuring high accuracy in training and testing while upholding privacy. The FL offers numerous advantages for conceptual learning in intelligent environments [4,5].

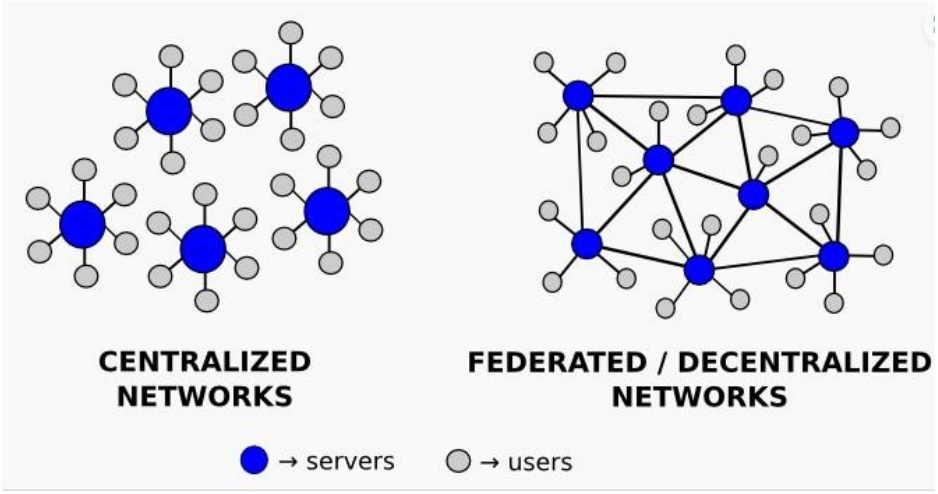

Fig.1. Comparison of Centralized and Decentralized Approaches

Table 1. Comparison of Centralized and Decentralized Approaches.

	Centralized Approaches	Decentralized Approaches
Security	Centralized systems can implement security measures such as firewalls, encryption, access controls, and intrusion detection systems.	Decentralized systems employ cryptographic techniques, consensus mechanisms, and distributed storage to enhance security.
Privacy	In a centralized model, individuals must trust the central authority to handle their personal data responsibly.	Decentralized models often prioritize privacy. Techniques such as pseudonymization, encryption, and privacy-enhancing protocols are employed to protect personal data.
Data Control	In a centralized model, the central authority has full control over the data, including access, storage, and processing. Individuals have limited control over their data and must rely on the central authority to enforce data protection measures.	Decentralized models empower individuals or entities with greater control over their data. Users can decide how and when their data is shared, granting explicit consent for each transaction. They have more autonomy and ownership over their data.
Single Point of Failure	Centralized systems have a single point of failure. If the central server or infrastructure fails, the entire system can become inaccessible or non-functional.	Decentralized systems distribute data across multiple nodes, eliminating the reliance on a single point of failure. Even if some nodes fail or go offline, the system remains operational, ensuring data availability.
Regulatory Compliance	In centralized arrangements, ensuring that they comply with data protection laws like GDPR can be challenging. centralized management of potential and data challenges in implementing data subject rights and consent management can pose compliance risks.	Decentralized models can facilitate compliance with data protection regulations. Users have more control over their data, transparent consent management can be implemented, and data minimization and pseudonymization techniques can be utilized.

1.1Background and Significance of Personal Data Protection and Privacy Control in the Digital Era

In the digital age, personal data has gained immense value and has become an integral component of numerous online services and transactions. Any information that can identify or locate a specific individual is considered personal data, including names, addresses, financial details, health records, and internet identifiers like cookies or IP addresses. The extensive gathering, storing, and processing of personal data, fuelled by the proliferation of digital technologies, social media, e-commerce, the internet, and Internet of Things (IoT) devices, pose significant challenges to data protection and privacy [6-9]. Enormous volumes of personal data are collected daily by companies for targeted advertising, personalized services, and understanding customer behaviour. However, concerns arise regarding potential abuse, illegal access, or disclosure of personal data due to this widespread data collection. Data breaches, where unauthorized parties gain access to private information, are becoming increasingly common, leading to identity theft, financial fraud, reputational harm, and other serious consequences for individuals. The constant monitoring and analysis of personal data give rise to ethical and societal issues, compromising individuals' privacy and autonomy. To address these concerns, governments and regulatory agencies worldwide have implemented data protection rules and regulations. The General Data Protection Regulation (GDPR) of the European Union is one of the most

comprehensive laws protecting personal information, emphasizing the rights of data subjects, the responsibilities of data controllers and processors, and the requirement for privacy by design in systems handling personal data. Consumer expectations play a crucial role in establishing trust between individuals and organizations regarding personal data protection. Consumers are increasingly concerned about how their data is handled, demanding transparency, control, and accountability from organizations. Organizations prioritizing data protection and privacy can build trust and foster stronger relationships with their customers [10-12].

Beyond legal and regulatory requirements, personal data protection raises ethical considerations. Respecting individuals' privacy and safeguarding their personal data is essential for upholding human rights, ensuring fairness, and promoting responsible data practices. The concept of centralized data management involves storing and controlling data in a single central location, typically governed by a single entity or organization. In this model, data is stored on centralized servers, and access to the data is managed by the central authority [13].

Table 2. Centralized and decentralized data management.

	Centralized Data Management	Decentralized Data Management
Architecture	In a centralized model, data is stored and managed in a single central location, typically controlled by a central authority or organization.	Decentralized data management distributes data across multiple nodes or participants in a network. Each node may have a copy of the entire dataset or a portion of it.
Data Control	The central authority has full control over the data, including access, storage, and processing.	Decentralized models empower individuals or entities with greater control over their data. Users have more autonomy in deciding how their data is shared, accessed, and used.
Data Security	Centralized systems typically implement security measures to protect data, but a breach or compromise of the central database can lead to widespread data loss or unauthorized access.	Decentralized systems employ cryptographic techniques, consensus mechanisms, and distributed storage to ensure data security and integrity. Tampering with data becomes more challenging due to the distributed nature of the system.
Scalability	Centralized systems may face scalability challenges as the volume of data and the number of users increases. The central server(s) can become a bottleneck, leading to slower performance.	Decentralized models can offer better scalability as data and processing are distributed across multiple nodes, allowing for parallelization and efficient resource utilization.
Privacy Concerns	Centralized models raise concerns about privacy as individuals must trust the central authority to handle their data responsibly.	Decentralized models often prioritize privacy. Techniques such as pseudonymization, encryption, and privacy-enhancing protocols are employed to protect personal data.
Single Point of Failure	Centralized systems have a single point of failure. If the central server or infrastructure fails, the entire system can become inaccessible or non-functional.	Decentralized systems are more resilient to failures or attacks since data is redundantly stored across multiple nodes, reducing the impact of a single point of failure.

1.2Decentralization

Decentralization involves distributing data and authority across a network of nodes or participants, eliminating the reliance on a single central authority. Blockchain technology is a prime example of a decentralized approach, where data is stored in a distributed ledger across multiple nodes. Each node maintains a copy of the entire blockchain, ensuring redundancy and data availability. Decentralized architectures provide several advantages. Decentralized networks employ cryptographic techniques and consensus mechanisms to ensure data integrity and security. The use of distributed ledgers makes it difficult for malicious actors to tamper with data [14].

1. Decentralized systems often employ techniques such as pseudonymization and encryption to protect personal data. Users have greater control over their data, and transactions can be conducted in a more privacy-preserving manner.
2. Decentralized systems enable individuals to have greater control over their personal data. Users can decide how and when their data is shared, granting explicit consent for each transaction.

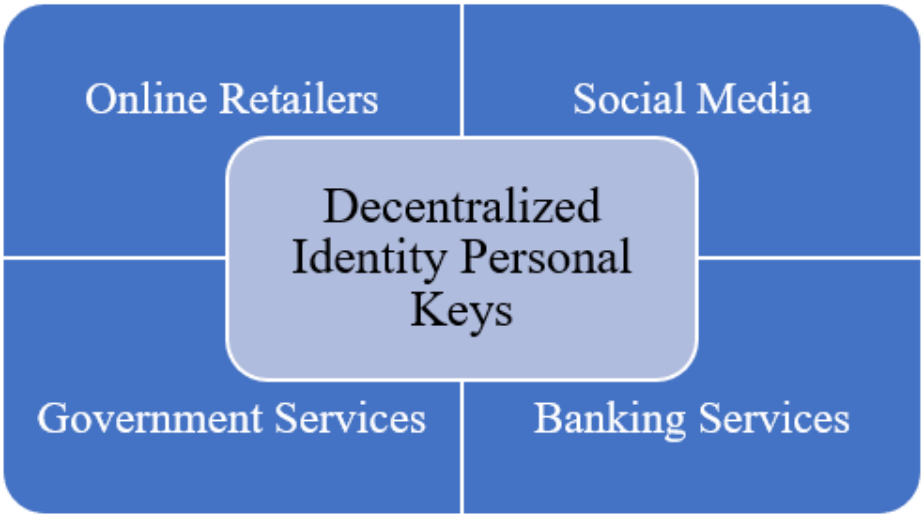

Fig. 2. Decentralized Identity

1.3Federated Approach

The federated approach involves collaborating entities or organizations that collectively manage and control data while preserving data privacy and security. In federated systems, data remains within the control of individual entities or organizations, and only aggregated or anonymized information is shared for analysis or collaborative purposes. Federated learning is an example of this approach, where machine learning models are trained across multiple distributed devices without the need to transfer raw data [15-17].

1. Federated systems enable data privacy by ensuring that sensitive data remains within the boundaries of each entity. Data can be processed and analysed collectively without exposing individual-level details.
2. Federated systems leverage the collective knowledge and expertise of multiple entities. Each entity contributes its unique data and insights, resulting in more comprehensive and accurate models or analyses.
3. By preserving data locality, federated approaches facilitate compliance with data protection regulations, especially when data cannot be easily shared due to legal, contractual, or privacy constraints. Federated systems distribute the computational and storage burden across multiple entities, reducing the reliance on a central infrastructure and optimizing resource utilization.

2 Decentralized and Federated Approaches as Potential Solutions

In response to the limitations and concerns associated with centralized data management decentralized and federated approaches have emerged as innovative solutions for enhancing data protection, privacy control, and overall data governance. These approaches distribute data across multiple entities and empower individuals with greater control over their personal information [18]. Centralized data management refers to a data management approach where all data is stored, controlled, and managed by a single central entity or authority. This centralized authority has complete control over data access, modification, and usage, which can offer certain benefits but also comes with various limitations. Both in the context of blockchain and AI, centralized data management has significant implications. In both blockchain and AI, decentralization is often considered a solution to address some of the limitations of centralized data management. Decentralized systems distribute data and control across multiple nodes or participants, providing benefits such as enhanced security, transparency, and reduced single points of failure. In blockchain, decentralized networks eliminate the need for a central authority and distribute the blockchain's data and consensus mechanisms among multiple nodes, ensuring transparency and censorship resistance. In AI, federated learning and edge computing are examples of decentralized approaches that enable training AI models across multiple devices or servers without centralizing all data in one location. While decentralization offers several advantages, it also comes with its own set of challenges, such as coordination among participants, potential for slower consensus, and added complexity. Striking the right balance between centralized and decentralized approaches is an ongoing area of research and development in both blockchain and AI technologies.

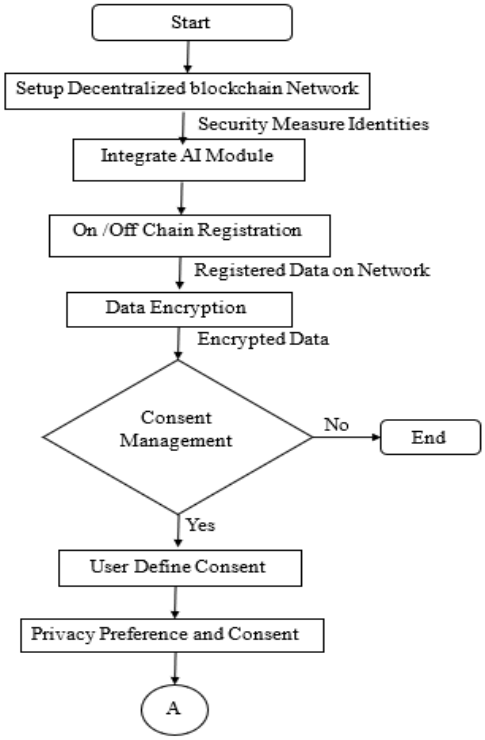

Fig.3. Flow chart of Decentralization and Federated Approach for Personal Data Protection and Privacy Control (Part A).

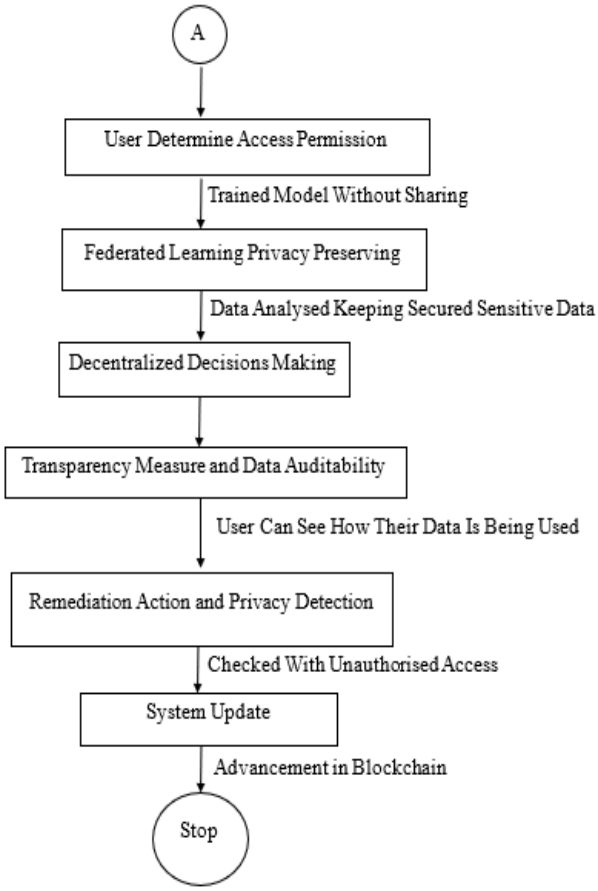

Fig.3. Flow chart of Decentralization and Federated Approach for Personal Data Protection and Privacy Control (Part B).

Both decentralized and federated approaches offer promising solutions for addressing the challenges of data protection, privacy control, and data governance in the digital era. By moving away from centralized models, these approaches prioritize security, privacy, and data autonomy, empowering individuals and organizations to have greater control over their data while fostering collaborative and privacy-preserving environments [19].

3 Related Work

Proposed A Blockchain Technology groundbreaking whitepaper on Bitcoin introduced blockchain technology, which forms the basis for decentralized ledgers. Blockchain's inherent security features and decentralized nature have been explored for various applications, including personal data management for instance, projects like "Self-Key" (Self Key Foundation) use blockchain to enable users to control their digital identity and personal data. Decentralized Identifiers (DIDs*)* are a key component of decentralized identity systems. DIDs, as defined by the W3C, enable verifiable, self-sovereign digital identities that individual's control [20].

Research on blockchain-based privacy protection has accelerated after 2020. The creation of a scalable and maintainable system using a blockchain-based Internet edge concept. Edge computing and blockchain can be used to build a privacy-preserving system that takes into account other limitations, such as energy costs. The topic of data sharing is tackled as a machine learning problem by combining privacy-preserving Federated Learning. Using blockchain technology, this develops a secure data exchange architecture for several remote parties. exchange of the data models rather than the actual data preserves data privacy. Federated Learning, a blockchain-based platform has begun developing applications for communication devices and the 5G network. Blocks are connected in chronological order by a distributed ledger structure that is exclusive to the blockchain [21].

All nodes in a decentralized system share and retain the saved data. In the majority of real-world settings, data, particularly personal data, is created and stored in separate locations, either on the devices of end users or in the data centres of service providers. Since most conventional ML algorithms operate in a centralized fashion, training data must be combined on a data server. From the perspective of anonymity [22].

A federated model is proposed to enhance the security of personal information, and its operation remains stable through the use of self-certification. To bolster user security and privacy in the Internet of Things, a secure authentication and key management strategy is recommended, supplementing the existing device identity authentication protocol for end devices and fog servers without certification. The complex tasks involved encompass collecting diverse data from multiple sources, aggregating, integrating, securely managing, and evaluating the data. Challenges arise not only from the high volume, speed, veracity, and heterogeneity of the data but also from factors such as industry competition, intricate administrative processes, and, notably, compliance with data protection laws and regulations like the EU General Data Protection Regulation when transporting data across organizations [23].

Furthermore, data owners share their local parameter modifications with a reliable third party, along with the training data. This study proposes a solution to address issues related to sparse data and examines the applicability of various modern models, presenting an alternative federated learning technique that reduces communication costs by transmitting fewer local parameters.

Federated learning is used in the distributed learning model created by to include several important activities. Federated learning was utilized to lower the hash rate by introducing recommendations for client selection in a resource-constrained mobile edge computing environment. This was included into the blockchain's consensus process. In fog computing, a learning-based intrusion detection system was created utilizing sample-selected extreme learning. Nevertheless, device-to-device communication and resource sharing were not taken into account by the analysis. To the best of our knowledge, no studies have examined how blockchain technology and federated learning might be used to efficiently safeguard communication privacy in a computing environment. Unauthorized access to a company's database results in a data breach since it provides hackers with sensitive personal information like passwords, account number, security numbers, and account details [24].

Identity theft and credit card number fraud, which can negatively affect a person's credit for a long time and often take months or years to rectify, were among the serious consequences of these documented breaches. The largest and most recent breach, believed to have been sponsored, affected billions of people when Yahoo's database was hacked in 2013–2014. While the aforementioned researchers have yielded the most recent findings and research outputs, there are still a number of distributed security vulnerabilities in federated learning that require fixing, including model extraction attacks and model reversal attacks. Data on names, email addresses, phone numbers, dates of birth, hashed passwords, and unencrypted responses to privacy inquiries were gathered by the hackers [25].

A superior centralized model while preserving the training data's dispersion among multiple clients. FL is beset with inconsistent client availability and subpar network performance. FL has been used over the years to protect sensitive personal data in a range of situations, such as video analysis, information inspection and

classification, and credit card fraud detection. Theoretical studies of convergence, network latency, and hostile attacks on FL are among the other hot issues at the moment. The centralized federated server has garnered more attention in the past few years. If at all possible, the concept of a server should be minimized. In order to properly accomplish convergence, the data would stay on edge devices and wouldn't require a central component like an aggregate server, according to a study [26].

Federated learning, characterized as a decentralized learning approach where local data remains with the data owner, stands out as a potentially valuable technique. This approach allows each data owner to possess a set of local learning parameters derived from the local model. Instead of merely providing the training data, data owners also share their local parameter updates with a trustworthy third party. The study scrutinized the appropriateness of several existing models and proposed a modification to address sparse data issues. Additionally, an alternative federated learning method was introduced, reducing communication expenses by transmitting fewer local parameters. A distributed learning model was formulated, considering various pertinent tasks through the implementation of federated learning [27].

The integration of federated learning into the blockchain consensus process was pursued to decrease the hash rate. An intrusion detection system based on learning was developed using sample-selected extreme learning in fog computing. However, this approach did not take into account resource sharing or device-to-device communication. To the best of our knowledge, no research has explored how blockchain and federated learning could collaborate to effectively enhance communication privacy in a fog computing environment. The study delves into Data Management from both centralized and decentralized perspectives [28].

To the best of our knowledge, certain advances provide security model sharing based on blockchain and federated learning with various protections to solve the issue of distributed security vulnerabilities.

4 Proposed Methodology

Algorithm for Decentralization and Federated Approach for Personal Data Protection and Privacy Control.

Step 1: Setup and Initialization

```
// Define the blockchain network
class BlockchainNetwork {
constructor (consensus Mechanism, security Measures) {
// Implementation details for blockchain network setup and initialization
}
// Other blockchain-related methods can be added here.
}
// Generate cryptographic key pairs for users
class User {
constructor() {
// Implementation details for generating unique cryptographic key pairs for each user
} }
// AI modules for privacy control, data encryption, access control, and personalized privacy
settings
class AIModule {
constructor() {
// Implementation details for AI modules responsible for privacy control, data encryption,
access control, and personalized privacy settings
} }
```

// Step 2: Data Registration and Encryption

```
// Helper function for data registration
function registerData(data, onChain) {
// Implementation details for data registration on-chain or off-chain based on &amp;#39;onChain&amp;#39;
flag
}
// Helper function for data encryption
function encryptData(data) {
// Implementation details for data encryption using techniques like homomorphic encryption
}
```

// Step 3: Consent Management and Access Control

```
// Define smart contract for consent management
class ConsentSmartContract {
constructor() {
// Implementation details for the smart contract handling user-defined consent rules
```

```
} }
// Helper function for access control
function enforceAccessControl(user, data, aiModule) {
// Implementation details for enforcing access control based on user-defined consent rules
and AI modules
}
// Step 4: AI-Driven Privacy Control
// Helper function for federated learning
function performFederatedLearning(models, data) {
// Implementation details for federated learning techniques
}
    // Helper function for privacy-preserving AI
function performPrivacyPreservingAI(data) {
   // Implementation details for privacy-preserving AI techniques like federated learning,
secure multi-party computation, etc.
}
// Helper function for personalized privacy settings
function generatePersonalizedPrivacySettings(user) {
// Implementation details for analyzing user preferences and generating personalized
privacy settings
}
// Step 5: Decentralized Governance
// Define smart contract for decentralized governance
class GovernanceSmartContract {
constructor () {
// Implementation details for the smart contract handling decentralized decision-making
and voting mechanisms
} }
// Step 6: Data Auditability and Transparency
// Helper function for data auditability
function verifyDataAuditability(data) {
// Implementation details for verifying data auditability using blockchain&#39;s transparency and immutability
}
// Helper function for transparency measures
function provideTransparencyInsights(user) {
// Implementation details for providing insights into data usage by AI algorithms and network participants
}
// Step 7: Privacy Violation Detection and Remediation
// Helper function for privacy violation detection
function detectPrivacyViolation(data) {
// Implementation details for using AI algorithms to monitor and detect potential privacy
Violations
}
// Helper function for remediation actions
function remediatePrivacyViolation(violation) {
// Implementation details for defining automated actions to address detected privacy violations
}
// Step 8: Continuous Improvement and Adaptation
// Helper function for machine learning feedback loop
function updatePrivacyControlMeasures(userFeedback) {
// Implementation details for enabling a feedback loop to adjust privacy control measures based on user
feedback
}
// Helper function for system updates
function updateSystem () {
// Implementation details for regular system updates to incorporate advancements and address vulnerabilities
```

It's important to note that the actual implementation of each step may involve utilizing various technologies, frameworks, and programming languages, depending on the specific requirements and available resources.

4.2Decentralized Personal Data Protection

Decentralization entails dispersing control, authority, and data across a network of nodes instead of relying on a singular central entity. Regarding personal data, decentralization entails distributing data storage, processing, and control among various entities, often involving users themselves, rather than centralizing all data in a single repository. The protection of personal data revolves around securing individuals' sensitive information, preventing unauthorized access, use, or disclosure. In an era characterized by extensive collection and processing of personal data, the safeguarding of this data has become of utmost importance [29].

Decentralization, when integrated with personal data protection, blockchain, and AI, establishes a robust framework for bolstering personal data privacy control. This amalgamation redefines how data is managed, accessed, and utilized. Decentralization, facilitated by blockchain technology, distributes data across a network of nodes, eradicating the vulnerabilities of centralized repositories. Users retain data ownership and dictate access permissions through cryptographic keys, fostering transparency and thwarting unauthorized access. Meanwhile, AI, empowered by the federated approach, enables collaborative model training exposing raw data. Techniques like homomorphic encryption ensure data privacy during analysis. This synergistic ecosystem is fortified by blockchain, which ensures secure, tamper-proof data transactions, access management, and identity verification. Users' consent choices and data interactions are immutably recorded, building trust and auditability. Consequently, this innovative integration empowers individuals to safeguard their personal data, participate in AI advancements, and maintain sovereignty over their privacy in the digital age.

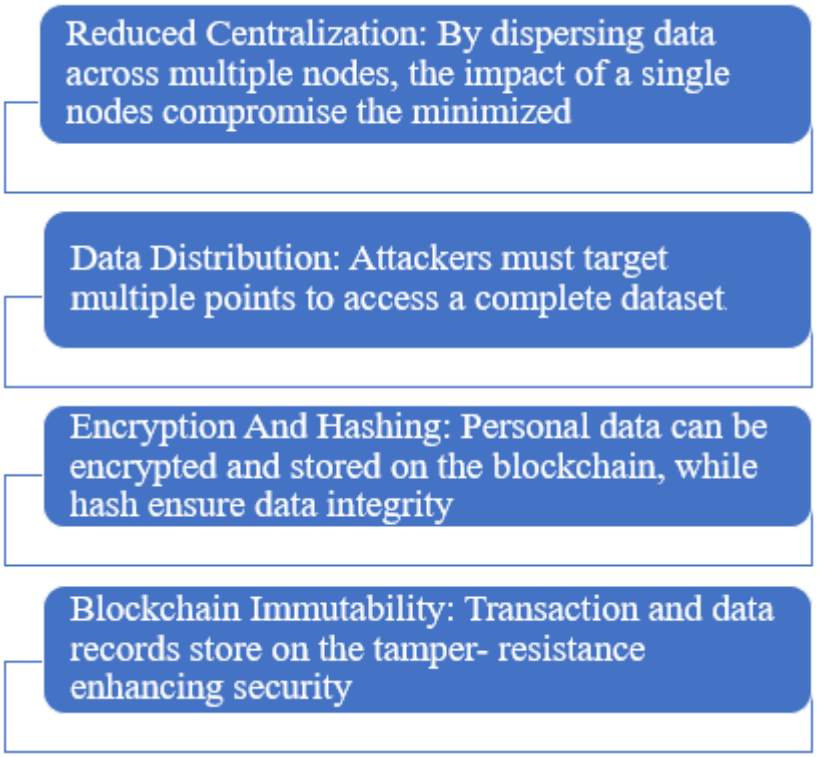

Fig. 4. Decentralized Personal Data Protection

Smart Contract for Data Access Control and Consent Management

To enhance control over data access and consent management, the incorporation of smart contracts is crucial. Smart contracts, defined as self-executing agreements with explicit code-based conditions, play a vital role. They offer automation and enforcement capabilities for access rules and user consent decisions, especially in the safeguarding of personal data. Smart contracts ensure that data transactions adhere to predefined requirements, eliminating the necessity for intermediaries or manual interventions by embedding these regulations directly into the code. By allowing users to articulate and oversee their data-sharing preferences transparently and automatically, smart contracts empower users to maintain authority over their data [30].

Consent Automation

Users can specify their consent preferences within a smart contract. For instance, they can dictate which data attributes can be accessed, by whom, and for what purposes. These preferences are programmed into the smart contract and executed automatically when data requests are made.

Granularity of Consent
Smart contracts enable users to express granular consent preferences. They can differentiate between different types of data usage and specify varying levels of access for different parties, ensuring that sensitive data remains protected.
Dynamic Updates
Users can update their consent preferences at any time, triggering automatic adjustments in the associated smart contract. This ensures that consent remains aligned with their evolving preferences.
Automated Enforcement
Smart contracts execute predefined rules autonomously. When a data request is made, the smart contract checks if the requested access aligns with the user's consent. If the conditions are met, access is granted; if not, the request is denied.
Transparency and Accountability
Smart contracts provide transparency by making consent preferences and data access rules explicit and visible on the blockchain. Users and authorized parties can audit and verify data interactions, enhancing accountability [31].

3.2. Federated Approach and Privacy Control

The federated approach is a cutting-edge method for cooperatively training machine learning models over a network of dispersed data sources while maintaining the locality of the data. It tackles the critical issue of performing data analysis without centralizing private information, making it extremely significant in scenarios that protect privacy. The federation protects individual privacy and complies with data protection laws by letting data to stay within the bounds of its source. A ground-breaking paradigm for machine learning called federated learning has evolved, giving priority to cooperative model training while protecting privacy. With this strategy, the difficulties of utilizing distributed data sources' collective intelligence while upholding data privacy are addressed. Federated learning achieves a careful balance between model accuracy and individual privacy by incorporating cutting-edge privacy-preserving techniques including homomorphic encryption, differential privacy, and secure multi-party computation [32-33].

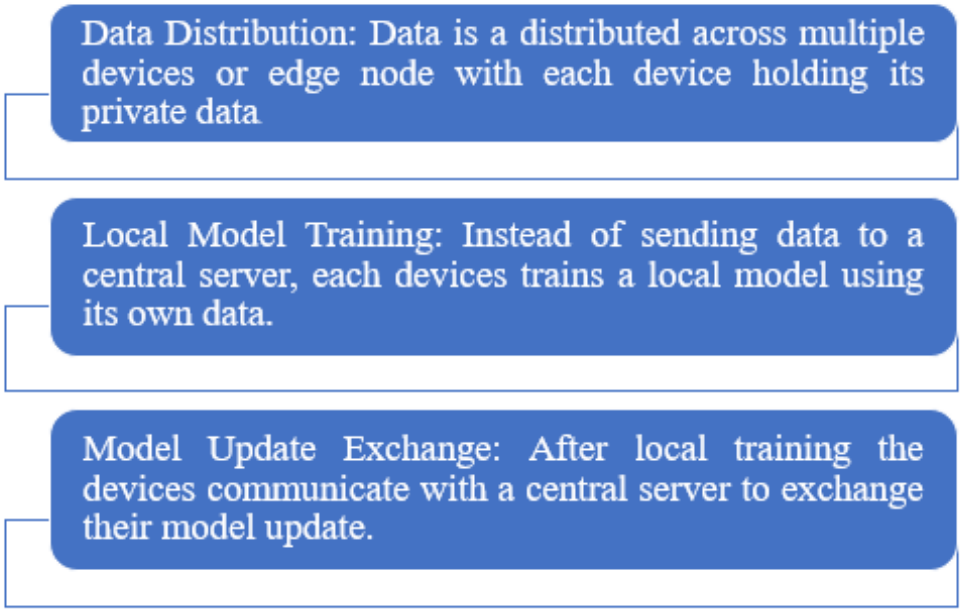

Fig. 5. Federated Learning Process for Privacy Control

Homomorphic cryptography
Without having to first decode the data, homomorphic encryption enables computation on encrypted data. This method makes it possible to compute on encrypted data while protecting it from outside parties, such as the model curator. Data owners can send encrypted model changes to the central server, which computes on the encrypted data, by using homomorphic encryption. The result is then encrypted by the server and sent back; only the data owners may decode it. This makes sure that private information is never made public while updating the model [34].
Distinctive privacy

To guarantee consistent outcomes in statistical queries on perturbed data without revealing specific details, differential privacy involves incorporating controlled noise into the data before analysis. In the context of federated learning, differential privacy can be applied by introducing noise to the aggregated model updates before transmitting them to the central server. This approach allows for meaningful analysis while thwarting any efforts to infer specific individual contributions from the model changes [35].

Secure Multi-Party Computation (SMPC)

The utilization of Secure Multi-Party Computation (SMPC) enables several participants to collaboratively perform computations on their individual private data without revealing specific data inputs. In the realm of federated learning, participants can create aggregated model updates without exposing their raw data. Each participant encrypts their update, and thereafter, the server carries out the computation of the aggregated model update without needing access to the individual updates. This ensures the preservation of privacy and confidentiality throughout the collaborative model training process [36].

Federated learning, enriched by privacy-preserving techniques like homomorphic encryption, differential privacy, and secure multi-party computation, revolutionizes collaborative model training while prioritizing individual privacy. This synergy empowers organizations to advance their models using collective insights while ensuring data security, privacy compliance, and trust among data contributors. As technology continues to evolve, this approach holds promise for reshaping the landscape of machine learning in a privacy-conscious manner [37].

3.3. Decentralized Identity Management

Self-sovereign identities (SSI) are a revolutionary concept in digital identity management that aligns seamlessly with the principles of decentralization. At its core, SSI empowers individuals with control over their own identity information, enabling them to manage, verify, and share their personal data without relying on centralized intermediaries. This concept reframes the traditional paradigm where organizations control and manage users' identities, often leading to data silos and privacy concerns. Instead, SSI emphasizes the individual's authority and agency over their identity, aligning perfectly with the principles of decentralization that distribute control and minimize reliance on central authorities.

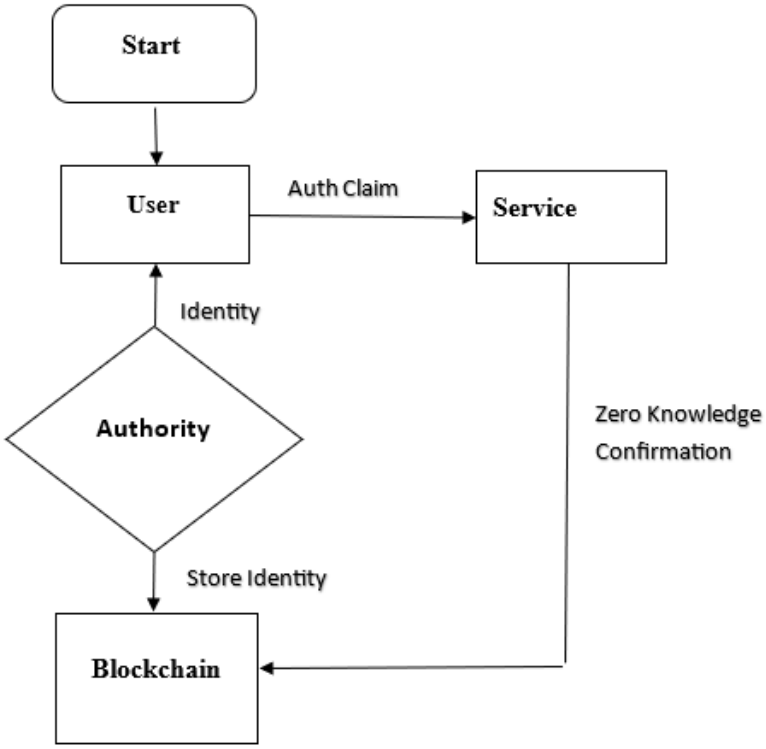

Fig. 6. Working of Self-sovereign identities (SSI)

Self-sovereign identities (SSI) are a revolutionary concept in digital identity management that aligns seamlessly with the principles of decentralization. At its core, SSI empowers individuals with control over their own identity information, enabling them to manage, verify, and share their personal data without relying on centralized

intermediaries. This concept reframes the traditional paradigm where organizations control and manage users' identities, often leading to data silos and privacy concerns. Instead, SSI emphasizes the individual's authority and agency over their identity, aligning perfectly with the principles of decentralization that distribute control and minimize reliance on central authorities. By using the SSI, you can connect your citizenship to your private key (your blockchain wallet). Following the issuing authority's endorsement of the veracity of your claim, third parties can use your public key to verify that the owner of the private key is in fact a citizen of the UK. You can make precise and otherwise anonymous requests to services by proving your identity claim in this manner [40]. By just stating that you are a British citizen, you can gain all rights and abilities that are exclusively based upon citizenship. There is no supporting evidence for that assertion (such as your passport, for example). Additionally, the service only obtains your public key in the process and nothing else.
Users typically share their public key so that service providers can verify claims with the issuing authority. The system can additionally ask the user to affirm that they agree for their address to be accessed for a specific claim by a certain company and then later revoke that consent. All of this occurs through the global blockchain network, giving users' private keys ownership over the data.

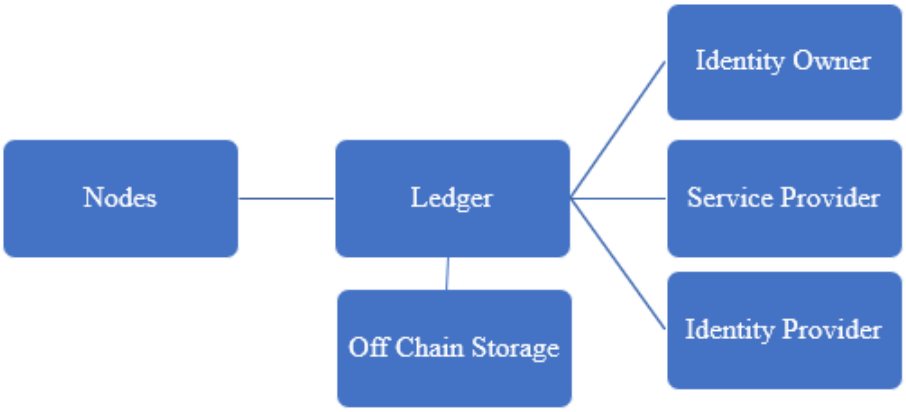

Fig. 7. Components of Blockchain-based-Sovereign Identity System

3.4. Impact of Legal Frameworks on Data Protection of Personally Identifiable Information (PII)

Legal frameworks such as the California Consumer Privacy Act (CCPA) in the United States and the General Data Protection Regulation (GDPR) in the European Union play a significant role in shaping data protection. The GDPR emphasizes individuals' rights to their personal data, mandating businesses to obtain explicit consent before utilizing it, establish comprehensive privacy policies, and grant individuals' access to, modification, or deletion of their data. Similarly, the CCPA empowers Californian consumers with control over their personal information, necessitating businesses to disclose data collection practices and provide opt-out mechanisms. Both regulations impose stringent penalties for non-compliance, compelling organizations to prioritize data privacy and security. While decentralized and federated approaches present promising solutions for data protection, they also prompt ethical considerations. The utilization of these technologies requires thorough evaluation to ensure alignment with ethical principles.

1. Which mandates that personal data must be handled properly, legally, and openly.
2. Which mandates that only valid, explicit, and precise purposes may be used to acquire personal data.
3. Which mandates that personal data must be sufficient, pertinent, and kept to a minimum.
4. Which mandates that accurate personal data be provided.

The rights of information subjects, the right to remain anonymous, and other rights must be protected in addition to the fundamental principles informed, Rights in respect to profiling and automated decision-making to uphold and confirm the accuracy of the GDPR policies. Federated learning demonstrates to be a method with very high potential as it is able to alleviate issues with machine learning such as the lack of data and the preservation of privacy from data from many sources. Creating a Federated Learning system with high predictive metrics while adhering to privacy norms and regulations will be one of the problems of the present and the future. Addressing these ethical considerations requires a holistic approach that involves collaboration between technology developers, policymakers, ethicists, and users. Balancing innovation with ethical

responsibility is key to realizing the full potential of decentralized and federated approaches while upholding individuals' rights and societal values. While decentralized and federated approaches offer promising solutions for data protection, they also raise ethical considerations. The use of these technologies requires careful evaluation to ensure that they align with ethical principles:

1. Transparency: Organizations must ensure that participants understand how their data is being used in federated systems and blockchain networks. Transparency is essential to maintain trust and informed consent.
2. Inclusivity: Decentralization and federation should not exacerbate existing digital divides or create barriers for marginalized communities. Efforts must be made to ensure equal access and participation.
3. Data Accuracy: The accuracy of decentralized and federated models relies on diverse and representative data. Ethical considerations include preventing biases and ensuring equitable representation across data sources.
4. Data Ownership: While decentralized approaches enhance data ownership, ensuring that participants have clear rights over their data and are appropriately compensated is crucial.
5. Consent and Control: Ethical use of these approaches necessitates respecting users consent and granting them control over data sharing. Participants should have the ability to revoke consent and access their data.
6. Security and Data Protection: Implementing strong security measures to stop data breaches and illegal access in decentralized and federated systems is an ethical consideration.
7. Interoperability: Ensuring that decentralized systems are interoperable and adhere to common standards is essential to prevent fragmentation and maintain open access.
8. Governance: Ethical questions arise around the governance of decentralized networks. Mechanisms for decision-making, dispute resolution, and evolving the network should be transparent and inclusive.

Addressing these ethical considerations requires a holistic approach that involves collaboration between technology developers, policymakers, ethicists, and users. Balancing innovation with ethical responsibility is key to realizing the full potential of decentralized and federated approaches while upholding individuals' rights and societal values.

4. Real-World Use Cases

Decentralized and federated approaches for personal data protection and privacy control have been applied in various real-world use cases. Here are some examples:

1. Healthcare: Federated learning enables collaboration among healthcare institutions while preserving patient privacy. Multiple hospitals can train a shared model on their respective local datasets without sharing sensitive patient information. This approach has been used for tasks such as disease prediction, medical imaging analysis, and drug discovery.
2. Finance: Financial institutions can leverage federated learning to collectively train models for fraud detection and risk assessment without sharing customer transaction data. By preserving the privacy of individual transactions, federated learning allows banks to collaborate while complying with data protection regulations.
3. Smart Cities: Decentralized and federated approaches are employed in smart city applications. For example, traffic management systems can analyse data from distributed sensors and cameras to optimize traffic flow without centralizing personal information or tracking individual vehicles. This enables efficient urban planning while respecting privacy.
4. Online Advertising: Federated learning has been explored in the field of online advertising to personalize ads without compromising user privacy. Advertisers can train models on local user devices to capture individual preferences while keeping personal data on the device. This approach helps deliver relevant ads without exposing sensitive information.
5. Education: Decentralized and federated approaches can enhance personalized learning and educational applications. By leveraging local device data, models can be trained to provide personalized recommendations, adaptive learning pathways, and student performance analysis while safeguarding student privacy [38].

These are just a few examples of how decentralized and federated approaches are being applied to protect personal data and privacy in various domains. As the technology evolves, we can expect more innovative use cases and applications leveraging these approaches to ensure privacy control while deriving insights and value from distributed data sources.

5. Challenges Encountered in Data Privacy and Protection.

Data privacy and protection laws has been considered an amazing achievement, however there exists a lot of challenges with the protection of user’s data.

1. Massive growth of data: Organisations are expected to adequately ensure protection of user’s personal information from any form of data breach and with the growing nature and size of data in our technologically

driven world and it becomes overwhelming to handle the billions of data. Also considering the various data privacy legislations and the complexity of its rules, there is a doubt that absolute compliance is even possible.
2. Conflict of laws: With the existence of several legislations on data privacy and protection, the possibility of conflict between several laws could often pose a lot of challenge to organisations whose users cut across various continents and affect their ability to adequately comply with each countries data privacy laws. For example, in relation to the definition of certain terms like data privacy, data subjects' consent, withdrawal of consent, right of erasure etc, which may often differ among the laws of several countries and this sometimes affect organisations decision as to determine what data privacy law to apply to their users.
3. Cost of maintaining data privacy: The cost of rectifying a data breach results in a huge loss of revenue to an organisation as such organisation faces intense regulatory penalties. Organisations are expected to adopt and take appropriate steps to ensure that adequate and specific security technologies are provided for, such as data backups and archiving in order for data to be safeguarded, recovered and restored.
4. Human Error complexities: Many data analysts have stated that human error is considered to be the biggest challenges in the data privacy for instance a lack of awareness by employees can significantly affect an organisations data privacy and protection system. This can occur through the use of weak passwords, falling for phishing scams etc and thus could result in user's data becoming vulnerable and lead to data loss.

6. Conclusion

Decentralization distributes data across multiple nodes or devices, reducing risk of single point of failure or unauthorized access. By storing data locally, individuals retain ownership and have greater control over its use. Additionally, decentralization reduces the dependency on centralized entities, such as data brokers or social media platforms, which often collect and monetize personal information without individuals' consent. The federated approach complements decentralization by enabling collaboration and analysis of distributed data while preserving privacy. Instead of centralizing data in one location, federated learning allows data to remain on local devices or servers while aggregating insights from various sources. This approach maintains data privacy by conducting computations locally and sharing only the model updates, thus protecting sensitive information. Implementing decentralized and federated systems requires technological advancements, such as secure peer-to-peer networks, encryption techniques, and privacy-preserving algorithms. Privacy-enhancing technologies like differential privacy can further strengthen these approaches by adding noise to data to prevent re-identification and protect against privacy breaches. Decentralized and federated approaches align with evolving data protection regulations, such as General Data Protection Regulation (GDPR). These regulations emphasize the need for individuals to have control over their data, consent to its use, and be informed about how it is processed. By adopting a decentralized and federated approach, organizations can demonstrate compliance with such regulations, fostering trust among users and reducing legal and reputational risks.

The decentralization and federated approach for personal data protection and privacy control offers a promising path towards a more privacy-preserving data ecosystem. By empowering individuals and enabling collaborative analysis while safeguarding their personal information, these approaches can strike a balance between data utility and privacy, fostering trust and innovation in the digital age.

References

1. Khalid, U., Asim, M., Baker, T., Hung, P.C.K., Tariq, M.A., Rafferty, L.: A decentralized lightweight blockchain-based authentication mechanism for IoT systems. Cluster Comput. 23(3), 2067–2087 (2020).
2. Nakamoto, S.: Re: bitcoin P2P e-cash paper. Cryptogr. Mail. List (2008)
3. Shen, H., Zhang, M., Wang, H., Guo, F., Susilo, W.: A cloudaided privacy-preserving multi-dimensional data comparison protocol. Inf. Sci. (Ny). 545, 739–752 (2021)
4. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: concept and applications. ACM Trans. Intell. Syst. Technol. 10(2), 1–19 (2019)
5. Wang, P., Wang, L., Leung, H., Zhang, G.: Super-resolution mapping based on spatial–spectral correlation for spectral imagery. IEEE Trans. Geosci. Remote Sens. 59(3), 2256–2268 (2020)
6. Zhou, W., Lv, Y., Lei, J., Yu, L.: Global and local-contrast guides content-aware fusion for RGB-D saliency prediction. IEEE Trans. Syst. Man Cybern. Syst. (2019)
7. J. Konecný et al., "Federated Learning: Strategies for Improving Communication Efficiency," CoRR, vol. abs/1610.05492, 2016, available: http://arxiv.org/ abs/1610.05492
8. S. Zhou et al., "PIRATE: A Blockchain-Based Secure Framework of Distributed Machine Learning in 5G Networks," CoRR, vol. abs/1912.07860, 2019, available: http://arxiv. org/abs/1912.07860.
9. Khan, L.U., Saad, W., Han, Z., Hossain, E., Hong, C.S.: Federated learning for internet of things: Recent advances, taxonomy, and open challenges. IEEE Commun. Surv. Tutorials (2021).
10. X. Chen., "When Machine Learning Meets Blockchain: A Decentralized, Privacy-Preserving and Secure Design," Proc. 2018 IEEE Int'l. Conf. Big Data (Big Data), (2018).

11. I. Hegedüs., G. Danner., and M. Jelasity.,: “Gossip Learning as a Decentralized Alternative to Federated Learning,” Proc. 2019 Int’l. Conf. Distributed Applications and Interoperable Systems — 19th IFIP WG 6.1, DAIS, (2019).
12. T. Li., A. K. Sahu., A. Talwalkar., and V. Smith., “Federated learning: Challenges, methods, and future directions,” IEEE Signal Processing Magazine, vol. 37, no. (2020).
13. P. Kairouz., H. B. McMahan., B. Avent., “Advances and open problems in federated learning,” (2019) [Online].
14. K. Gai., Y. Wu., L. Zhu, Z., Zhang, and M. Qiu., Differential privacybased blockchain for industrial internet-of-things, IEEE Transactions on Industrial Informatics (2020)
15. H. Kim., J. Park., M. Bennis., and S. Kim., Block chained on-device federated learning,” IEEE Communications Letters (2020).
16. Y. Liu., J. Peng., J. Kang., A. M. Iliyasu., D. Niyato., and A. A. A. ElLatif., A secure federated learning framework for 5g networks, IEEE Wireless Communications (2020).
17. L. Ma., Q. Pei., Y. Qu., K. Fan., and X. Lai., Decentralized privacy preserving reputation management for mobile crowdsensing, in International Conference on Security and Privacy in Communication Systems. Springer (2019).
18. Y. Jiao., P. Wang., D. Niyato., and K. Suankaewmanee., Auction mechanisms in cloud/fog computing resource allocation for public blockchain networks, IEEE Trans. Parallel Distribution. System (2019).
19. X. Wang., Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, “In-edge ai: Intelligentizing mobile edge computing, caching and communication by federated learning,” IEEE Network, (2019).
20. V. Smith., C. Chiang., M. Sanjabi., and A. S. Talwalkar., Federated multi-task learning, in Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems (2017).
21. J. Konecny., H. B. McMahan., D. Ramage., and P. Richt., ´ arik, Federated optimization: Distributed machine learning for on-device intelligence, CoRR, (2016).
22. X. Wang., Y. Han., C. Wang., Q. Zhao., X. Chen., and M. Chen., In-edge ai: Intelligentizing mobile edge computing, caching and communication by federated learning, (2019).
23. V. Smith., C. Chiang., M. Sanjabi., and A. S. Talwalkar., Federated multi-task learning, in Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems (2017)
24. J. Konecny., H. B. McMahan., F. X. Yu., P. Richt., A. T. Suresh., and D. Bacon, Federated learning: Strategies for improving communication efficiency, (2016).
25. T. Nishio., and R. Yonetani., Client selection for federated learning with heterogeneous resources in mobile edge in 2019 IEEE International Conference on Communications, (2019).
26. Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, “Blockchain and federated learning for privacy-preserved data sharing in industrial iot,” IEEE Transactions on Industrial Informatics, (2019).
27. Yoon J., Democratic Senators Introduce the Consumer Online Privacy Rights Act. [Internet]. (2019).
28. Aktypi A., Nurse J.R.C., Goldsmith M. Unwinding ariadne’s identity thread: Privacy risks with fitness trackers and online social networks international workshop on multimedia privacy and security at ACM conference on computer and communications security (2017).
29. Mehmood A., Natgunanathan I., Xiang Y., Hua G., Guo S.Protection of big data privacy IEEE Access, 4 (2016).
30. Belen Saglam R., Nurse J.R.C. Is your chatbot GDPR compliant? Open issues in agent design international conference on conversational user interfaces, ACM (2020).
31. Ramachandran G.S., Radhakrishnan R., Krishnamachari B. Towards a decentralized data marketplace for smart cities E international smart cities conference, IEEE (2018).
32. Lönnfält I., Sandqvist J.Blockchains, the new fashion in supply chains?-the compatibility of blockchain configurations in supply chain management in the fast fashion industry Master’s thesis Gothenburg University (2018).
33. Rumbold J.M., Pierscionek B.K., What are data? A categorization of the data sensitivity spectrum Big Data Res, 12 (2018).
34. Purtova N., The law of everything. broad concept of personal data and future of EU data protection law, law Innov Technology, (2018).
35. Singh Suyogita., Verma Satya., Federated Learning for GDPR. The journal of Machine Intelligence Research Lab (MIR lab), (2013).
36. Satya Bhushan Verma., Abhay Kumar Yadav., Detection of Hard Exudates in Retinopathy Images ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal Regular Issue, Vol. 8 N. 4, 41-48 eISSN: 2255-2863 (2019)
37. Satya B Verma., Shashi B V., Data Transmission in BPEL (Business Process Execution Language), ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal Regular Issue, Vol. 9 N. 3 105-117 eISSN: 2255-2863 (2020).

38. SB Verma., Brijesh P., and BK Gupta., Containerization and its Architectures: A Study, ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, Vol. 11 N. 4 , 395-409, eISSN: 2255-2863, (2022).
39. Verma, S. B., & Saravanan., C. Performance analysis of various fusion methods in multimodal biometric. In 2018 International Conference on Computational and Characterization Techniques in Engineering & Sciences (CCTES) (pp. 5–8). IEEE. (2018).
40. Verma, S.B., Yadav, A.K. Hard Exudates Detection: A Review., Emerging Technologies in Data Mining and Information Security. Advances in Intelligent Systems and Computing, vol 1286. Springer, Singapore (2021).