JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Towards Improved Privacy in AI and Machine Learning Applications: Challenges and way forward.

Ehigiator Iyobor

Egho-Promise Department of Computer Science
University College Birmingham, Birmingham, United Kingdom

George Asante
Department of Information Technology Education, Akenten Appiah-Menka University of Skills Training and Entrepreneurial Development, Kumasi, Ashanti Region, Ghana

Hewa Balisane
Business School, The University of Law, Manchester, United Kingdom

Folayo Aina
Department of Computing, School of Engineering and Computing, University of Central Lancashire, Preston, United Kingdom

Halima Kure
Department of Computer Science and Digital Technologies, University of East London, London, United Kingdom

Abstract

Artificial intelligence (AI) and machine learning (ML) systems depend heavily on large datasets to function effectively. These datasets contain details of people and can be names, addresses, account numbers, credit card numbers, health data, and behaviour data, among others. Such enormous amounts of data are typically collected, stored, and analysed. This could lead to privacy violations if not sensitively done. Privacy violations are caused by causes such as inadequate security, illegal access, or hacking, all of which have negative repercussions for the individuals and businesses involved. This study aims to identify the privacy concerns unique to AI and ML applications and assess the efficacy of different privacy-preserving approaches. Specifically, the study seeks to identify the privacy challenges to AI and ML applications and evaluate the effectiveness of various privacy-preserving techniques that apply to AI and ML applications. This study used a qualitative research approach based on case studies, and data was acquired from secondary sources such as published papers, websites, and publications. It has been found that recent advances in

artificial intelligence (AI) and machine learning (ML) have shown a new dawn in corporate functions, guiding efficiency, innovations, and insights across several industries. Although the use of personal data by these technologies has been extensively adopted, it has raised several privacy issues. The research identified certain privacy issues specific to AI and ML applications, such as overfitting, data leakage, illegal access, model inversion attacks, re-identification, and Privacy audits. It can be concluded that, despite the continuous development of AI and ML technologies and their successful deployment in all fields of human activity, privacy remains one of the most pressing concerns that are yet to be adequately addressed. Designing AI and ML applications to achieve superior levels of performance while maintaining individual privacy is a complex task that will require a combination of technical, normative, and ethical approaches, as well as the collaborative efforts of technical experts, ethicists, legislators, and users. Techniques such as Data Anonymisation and Pseudonymisation, differential privacy, federated learning, homomorphic encryption and secure multi-party computation should be used to improve privacy in AI and ML applications.

Keywords: Privacy, AI, Machine Learning, Privacy-Preserving techniques

1. Introduction

Artificial intelligence (AI) and machine learning (ML) systems depend heavily on large datasets to function effectively [1]. Such datasets contain details of people and can be names, addresses, account numbers, credit card numbers, health data, and behaviour data, among others. Such enormous amounts of data are typically collected, stored, and analysed. This could lead to privacy violations if not sensitively done. Sharma and Oriaku [2] opined that privacy violations result from factors such as poor security measures, access of unauthorised persons or hacking, and these have undesirable consequences for the individuals and firms concerned. The historical point of view on privacy is considered as a representative that evolves by relying on technologies. At the initial stage, the privacy concerns focused on securing physical access to information in order to enable any unauthorised personalities to access it. However, with the advent of digital technologies and the internet, privacy concerns have changed dramatically [3] In the early part of the twentieth century, privacy was described as the 'right to be let alone'; it meant that privacy equals the justified avoidance of any prior documented infringement on the subject. This notion paved the way for drafting many of today's privacy laws and statutes.

The Ribeiro-Navarrete et al [4] research found that the changes in methods of handling technology in today's world and an individual right to privacy have become the major set-up. In the present-day context, it is differentiated by the application of refined technologies such as AI and ML; however, the advancement of comparatively protected regulations has been steadily slow. This lag sometimes results in either miscreants or tech companies incorporating new technologies that are contrary to the rights of persons' privacy and liberty. AI and ML systems can be described effectively in terms of this dynamic because the creation and use of these technologies have progressed and multiplied, engendering novel privacy threats that are not promptly addressed by regulations. The purpose of this study was to identify the privacy concerns unique to AI and ML applications and assess the efficacy of different privacy-preserving approaches. Specifically, the study seeks to identify the privacy challenges that are specific to AI and ML applications. Furthermore, the study seeks to evaluate the effectiveness of various privacy-preserving techniques that are applicable to AI and ML applications.

2. Privacy Challenges in AI and ML

2.1. Data Collection

Gathering data is the most crucial stage of introducing AI and ML systems into practice and creating AI-associated applications [5]. This phase entails assembling a massive amount of data from various sources to train the models to identify patterns and structures and to deploy algorithms to predict the outcomes of certain processes and make decisions. However, some concerns are raised: the violation of people’s rights to Privacy and ensuring ethical use of their data. Major issues in the data collection phase include the consent process, data minimisation and Transparency.

2.1.1. Informed Consent

In the collection of data, probably the most significant privacy issue is inadequate consent from the data subjects. Informed consent entails the knowledge that an individual requires when data is being collected, where it is being used, who is going to use it, and what the dangers of participating in data sharing are [6]. This process should not be an empty ritual; it has to be clear, understandable and exhaustive to provide necessary information for a person or group on their data. However, Privacy has proved to be hard to maintain in the utilisation of AI and ML, especially when data is used in complex and sometimes obscure ways; thus, the issue of informed consent proves to be a major problem [7]. As a result, users must be educated about the ramifications of their data being used to train models that would impact decisions in a number of fields, including medical, financial, and policing.

2.1.2. Data Minimisation

Data minimisation is one of the data protection principles that is crucial for the intended purpose. This is important in mitigating the privacy risks, as excessive data collection not only increases the misuse potential but also amplifies the data breach consequences. In the context of AI and ML, the tendency to improve the model's accuracy must be linked to the need to protect the Privacy of individuals [8]. Data minimisation could be viewed as an activity that is based on significant planning and a clear understanding of specific data needs for AI/ML tasks. This involves aggregating and anonymising data to reduce the amount of personal information that is collected and stored.

2.1.3. Transparency

According to Androniceanu [9], in the data collection process, transparency is considered an important aspect to build trust and ensure accountability. Regarding personal rights, users should be in a position to know how specifically their data will be utilised, who will use it or for what purpose the data is collected. This should be a full cycle of transparency covering the entire data collection process and the method of using the data. For instance, should inform the affected individuals when it is realised that the intended use of the data has changed or when the identity of the third parties to whom the data will be transferred changes. Transparent practices not only strongly contribute to the user's control over the personal information necessary for further consent but also help to avoid situations when the user does not really know who is using their information and for what purposes [6].

2.2. Data Storage and Processing

Data storage and processing phases in artificial intelligence (AI) and machine learning come out as the most significant phases that raise massive privacy risks and must be handled discreetly to protect sensitive information [10]. These phases include safeguarding big data and computation essential for the training and deployment of AI solutions. The key security concerns include ensuring robust security, integrating strict access control, forming clear data retention policies, and maintaining integrity [11]. These issues must be resolved to safeguard personal information, reconstruct the clients' trust, and conform to the legal requirements.

2.2.1. Security

Security is considered a crucial aspect of the storage and processing of data. Data security, therefore, involves the application of good security standards to protect the data from breaches, unauthorised access, and other cyber-attacks. Integrating encryption is considered an important strategy to enhance data security [12]. Where the data is stored and when the data needs to be transferred from one place to another, it should be encrypted. However, to ensure that data remains confidential, advanced encryption standards (AES) and secure communication protocols such as Transport Layer Security (TLS) can be integrated. Moreover, Al-Matari et al. [13] highlighted that potential safety risks can be mitigated by implementing regular security audits, vulnerability assessments, and penetration testing. To monitor and respond to security incidents promptly, implementing intrusion detection systems (IDS) and intrusion prevention systems (IPS) plays an essential role. Strict measures are considered crucial to ensure that only authorised personnel can access data [14] Implementation of role-based access control (RBAC) in organisations makes it possible to operate under the principle known as least privilege, where clients get access only to data that is relevant to their job. It assists

in providing permitted access to the data based on the duties of the people in a particular group/department. However, multi-factor authentication (MFA) provides a sense of security by requiring users to prove at least two aspects before accessing the system [15].

2.2.2. Data Retention

Data retention policies are considered crucial for managing how long data is stored and ensuring its deletion when no longer needed. According to Silva et al [16], retaining data poses a significant privacy risk, as the longer data is stored, the greater the chance of it being exposed in a breach or becoming inaccurate. However, to mitigate such risk, it is important to establish clear data retention policies [17]. To cope with this situation, organisations must determine the period for retaining data that is demanded by the law, regulations, and business discretion. For the data deletion when the retention time has elapsed, an automated mechanism must be implemented. This not only saves the users' data privacy but also follows rules such as the General Data Protection Regulation (GDPR), which lets data be stored only for a certain time frame.

2.2.3. Data Integrity

According to Fizur [18], data integrity is considered important to ensure that data remains consistent, accurate, and unfiltered during the process of processing and storage. However, when changes are affected by the wrong personnel, data is corrupted, or when systems break down, the integrity of the data is compromised. With this, the checksums, together with the cryptographic hash, can be of main help when it comes to making someone realise that a given data has changed. Moreover, regular activities like data checks, data auditing, and other consistency-checking methods can also be conducted to avoid data corruption [19]. On the same note, version control and copying the data can also help in enhancing the possibility of data recovery in transit or any time that it gets corrupted. Data reliability must be addressed when using AI and ML, where the reliability and accuracy of the data directly impact the trustworthiness and performance of the models.

2.3. Model Training and Inference

Model training and inference are vital in the artificial intelligence (AI) and machine learning (ML) system development phases [20]. In these stages, models learn from data and make predictions on new input data. However, these processes can pose serious problems with respect to Privacy, as the information is disclosed and can be misused. There are typical challenges that are as follows; overfitting, membership inference attacks, model inversion attacks, and differential Privacy.

2.3.1. Overfitting

Overfitting usually happens when the model memorises information that is not only present in the training set but also the noise information included in the training dataset [21]. This can result in a model that is well-suited to classified data but performs poorly in new data sets. Far more seriously, overfitting leads to the disclosure of information that is included in the training set. For instance, if a model gets trained on specific data points of certain people when probed, it is likely to bring out details about such people. To reduce overfitting, some of the methods that can be used are regularisation, cross-validation, and dropout [22]. These methods facilitate the ability of the model to perform better when exposed to unseen data while avoiding aspects of memorising data that violate Privacy.

2.3.2. Membership Inference Attacks

Membership inference attacks target to find out whether the specific data point was used in the training of a model [23]. This type of attack can be very dangerous in terms of Privacy, especially if some data included in the training set is sensitive. For example, information about an individual provided for training in a medical diagnosis model can be used to suggest that the individual had a given medical condition. To counter such attacks, methods include adding noise to the training process, simplifying the models, and applying differential privacy work [24]. As a result, these approaches reduce the difference between the training and non-training data sets, thereby minimising the exposure of individuals' data.

2.3.3. Model Inversion Attacks

The model inversion assaults look to construct the normal structure of a trained model to examine the input data [25]. It involves the use of the model-learned parameters to calculate some of the attributes of the input

data that should otherwise not be disclosed. For example, a new attacker can utilise a facial identification model to estimate the image of an individual's face based on the result of the model. However, to prevent model inversion attacks, the former has to make sure that the number of information bits in the model output is held to a minimum. Some of the approaches are output perturbation, where some noise can be introduced to the model, and restricting information input to the model [26]. Moreover, it is also possible to train the model and make it simply invisible to the attacker even if they try to perform the inversion attack. This will increase the privacy aspect of the model.

2.4. Identifiable Information and Re-identification Risks

According to Liu et al[27], in most cases, all artificial intelligence (AI) and machine learning (ML) handle different types of identifiable information; therefore, Privacy becomes a significant factor. Identifiable information includes information that, one way or another, will inform an intruder about the owner of the data. The number of risks linked with handling such kind of data involve potential re-identification threats, linkage attacks, and other similar violations. However, as per Di Minin et al [28], the discussed risks may be prevented through the reduction of data identifiers, maintaining a database that does not generate links to specific users, non-combination of the particular databases, and the regular check for Privacy.

2.4.1. Re-identification

Martinez and Herrera [29] highlighted that despite the efforts to anonymise data, difficulties like the possibility of re-identification persist. Re-identification usually occurs when certain data is de-identified and then reprocessed together with another data set. It may be even refined in a way to get to the identity of the individuals. This risk is rather high in the era of big data that enables the merging of several sets of data that can be easily accessed. Therefore, to eliminate the risks of re-identification, it is important to implement accurate anonymisation methods and evaluate their effectiveness [30]. The use of specific techniques like k-anonymity, l-diversity, and t-closeness play an important role in solving privacy problems by ensuring that individuals cannot be identified from the anonymised data [31]. For example, if k-anonymity provides that for each record in the database, one could not differentiate it from at least k-1 other records in the same database.

2.4.2. Linkage Attacks

One other enormous threat type is the linkage attacks, where the attackers compile numerous sets to learn about particular persons [32]. These attacks exploit correlation with other databases, for instance, a health record of a patient is linked to the social account of the patient with a view of accessing their health records. Some of the general standards of data management provide defence or protection that counterpoints linkage attacks. This involves limiting data access to a few individuals, applying differential analysis and computation measures, and updating anonymisation techniques. However, organisations should conduct a thorough risk assessment to understand the potential linkages and establish adequate measures [33]. For instance, in differential Privacy, some randomness is added to the data so that the attacker will not be able to link the datasets.

2.4.3. Privacy Audits

Ferra et al [34] opined that regular privacy audits are important in meeting the set expectations in the privacy standards and identifying some threats. Privacy audits are the processes of assessing the treatment of the information, security measures, and compliance with the legislation and norms. They may be applied to identify areas in an organisation's privacy program that must be addressed. Privacy impact assessments (PIAs) are considered a proactive approach that enables the evaluation of risks that accompany data processing to individual's rights to Privacy and, where possible, mitigation of such risks [35]. However, La Torre et al[36] argued that privacy audits should be conducted in such a way that they are frequent processes, ensuring that the practice of Privacy is in response to new risks and regulatory changes.

Therefore, Murakami and Takahashi [37] concluded that when managing identifiable information in the AI and ML systems, one has to secure the information in ways that eliminate the risks characteristic of re-identification. This implies that data anonymisation and pseudonymisation provide basic protection while preventing linkage attacks, and re-identification requires efficient and progressive measures. In addition,

privacy audits also play a crucial role in maintaining the levels of Privacy and reveal the emergence of new threats [36]. Thus, by adapting these methods, organisations can effectively protect individual Privacy, maintain dependability, and observe compliance with regulatory requirements.

3. Techniques for Ensuring Privacy in AI and ML

3.1. Data Anonymization and Pseudonymization:

Data anonymisation and pseudonymisation are considered crucial techniques to ensure privacy in AI and machine learning [38]. Majeed and Hwang [37] defined data anonymisation as the process of transforming the data in order to obscure the identity of persons involved in the data either directly or indirectly. Some of the most frequently used methods are generalising, which consists of replacing original data with more general data, and suppression, which involves leading out data from a set of data. Another method is masking. For instance, the substitution of the exact birth dates with the age can be a method of data masking applied for anonymisation [39]. The main goal is to achieve the outcome where even if data is stripped of identifiers and merged with other databases, identifiers can never be worked backwards to reveal people. In pseudonymisation, personal identifiers are replaced by artificial names or codes. While pseudonymised data can be re-identified if other information is obtained, it is advantageous in that it does not contain actual identifiers in the dataset.

Techniques like data masking, randomisation, and k-anonymity make sure that within the dataset, no one record can be differentiated from at least ‘k-1’ other records. However, pseudonymisation is presented as the opposite of anonymisation, as it enables reversible changes with necessary control, especially in cases of regulations that require the identification of the specific individual.

In addition, Zuo et al [40] highlighted that data anonymisation is significant in the context of pervasive healthcare, where techniques such as differential privacy and noise addition play a crucial role in ensuring that patient data is protected while still making sense of it. Thus, the proposed work highlights the challenges of maintaining utility while privacy is still a concern. However, pseudonymisation in the healthcare sector allows one to keep the link between the data and patients’ records without compromising their identities. This method helps track outcomes over some time and patient management besides guarding and entrusting personal details. Moreover, Majeed et al [38] study concentrated on the clustering-based anonymisation techniques, where data points are first clustered and then anonymised in their clusters in order to prevent re-identification. This method assists in preserving the functionality of the data used for analysis to enhance the protection of individuals’ privacy. Pseudonymisation is carried as one of the ways through which privacy may be attained in datasets that require some level of linkability. Majeed et al [38] insisted on the need for secure key management practices to reduce unauthorised re-identification risks.

3.2. Differential Privacy

Differential privacy constitutes one of the foundations in the quest for the technique to reconcile the immense utility of value from the analysis of large datasets without intruding on the rights of data subjects. It is based on the concept of adding noise to the data or query results, thereby obscuring the individual contributions while enabling meaningful statistical analysis. The In particular, in the case of federated learning, there are privacy issues, so applying some of the methods, such as differential privacy, becomes essential. Therefore, if these measures are used, one can proceed with the process of federated learning without violating the subject’s rights to privacy of data. Similarly, according to Silva et al [16], even under the umbrella of other AI, differential privacy can be used to enhance privacy. It involves delving into numerous strategies and methodologies for implementing differential privacy, considering the intricacies of diverse AI techniques.

According to the use of differential Privacy in the training of the model, it is possible to safeguard against three different genres of attacks, such as membership inference and model inversion [41]. It also provides a reasonable guarantee of users’ Privacy, and it allows the necessary information to be extracted from the data without the loss of overall applicability of the model.

Additionally, Padmanaban [42] contributed by exploring privacy-preserving architectures, advocating for differential privacy as a fundamental concept. The purpose of such architectures is to provide the highest level of confidentiality regarding the data and make the information quite useful for the intended analysis.

3.3. Federated Learning

Federated learning appears as one of the first solutions to the complex problem of dealing with both the usefulness of the data and privacy concerns in the context of ML and AI. In its essence, this revolutionary approach, explained by [43] embodies a revolution in the model training process through the use of interconnected devices or servers. This way, federated learning reduces the need to bring sensitive data to a central place, which means a tremendous decrease in potential risks connected to the violation of privacy rights. Possibilities of privacy leakage make the federated learning protection concept highly effective due to the decentralisation of computation and model updates [44].

Like other research works, Li et al [45] also stress data privacy preservation is a paramount issue where federated learning emerges as an implemental and viable solution. Since, by its inherent mechanisms, the federated learning approaches do not share the individual data with which the model is trained, with a central server or any other external authority. This effectively mitigates the issues of privacy linked with centralized data processing architectures. Moreover, Jagarlamudi et al [74] explored the subtle aspects of privacy quantification in the context of federated learning. They emphasised that the methodology's enhancement and development are ongoing in order to strengthen privacy protection in different ecosystems. However, due to the strict privacy measurement frameworks, the federated learning methodologies are constantly evaluating and enhancing the privacy-preserving measures, which make the methods more stable protectors of the individual's privacy rights.

3.4. Homomorphic Encryption

Homomorphic encryption emerges as one of the modern cryptographic techniques that is considered promising for revolutionising confidentiality protection in the data processing field. Asante et al [46] indicated that, with homomorphic encryption, you can perform computations on the encrypted data without the need for the decryption of the information first. This particular characteristic implies that data to be safeguarded when computation is underway remains protected from the time the data is input to the time it is output. By using homomorphic encryption technology as an AI algorithm improvement, Wang et al [47] were able to uncover the alliance between homomorphic encryption technology and AI in the advancement of privacy in AI, such as IoT. Thus, homomorphic encryption prevents leakage of deep learning data by computing on homomorphic encrypted data framework within the AI.

In addition, Rahman et al [48] reemphasised the significance of heteromorphic and homomorphic encryption with regard to the edge networks for the execution of privacy-preserving AI composition frameworks. The homomorphic encryption techniques resulted in empowering edge devices to implement efficient AI operations on the encrypted data at the periphery, preserving data confidentiality. Moreover, Yaji et al [49] proposed that integrating homomorphic encryption with other emerging technologies, such as blockchain, further improves the versatility and applicability across the different AI applications.

3.5. Secure Multi-Party Computation (SMPC):

The next promising concept regarding the attainment of privacy and cooperative computation by several parties is the Secure Multi-Party Computation (SMPC). According to Skarkala et al [50], SMPC is a solution that facilitates function computation through the cooperation of several parties without revealing other parties' inputs. Most importantly, participants in an SMPC protocol receive only the result of the operation performed but not the inputs provided by other users. This confidentiality of inputs is particularly useful where inputs are in areas that require keen anonymity, especially when the inputs are to be gathered or pooled by different entities, as noted by Schaller [51]. To reduce the chances of exposing the inputs provided by the members and maintain the privacy of all members involved, SMPC conceals the inputs of every member throughout the computation process.

Zhou et al [52] initially proposed secure computation protocols; this paved the way for the refinement and development of the SMPC techniques. These protocols allow two distinct parties to cooperatively compute the given data of two different datasets simultaneously without revealing the original data to each other, this helps in overcoming privacy issues and carrying out effective data processing and collaboration. Moreover, Pilton et al [53] highlighted that it is especially useful in cases where information is collected from various sources, and summative and comparative analyses need to be done without violating individuals' rights to

privacy. Secure Multi-Party Computation (SMPC) is, therefore, the first stepping stone for data privacy in data mining, including collaborative analysis. It offers a solid and mathematically credible workbench for efficient, secure computation in distributed environments. With the future developments of the field, SMPC is very promising for creating better privacy protections in sectors such as health, finance, and distributed computing systems where confidentiality must be maintained, but large-scale computation must also occur.

3.6. Adversarial Training

Adversarial Training is a fundamental strategy to enhance the robustness of machine learning models against adversarial threats and consequently maximise the assurance of information privacy. Adversarial examples, as mentioned by Goodfellow et al [54] are original images designed to exploit the vulnerabilities in particular types of learning models, potentially giving out specific information from the outputs. These attacks directly endanger the confidentiality of data, especially in cases where the data is processed with the use of machine learning algorithms, emphasising the need for robust defence mechanisms to protect it.

As per Madry et al [55]adversarial training can be defined as a form of accurate planning that helps to strengthen the model's immunity to these attacks. By adding adversarial examples to the perturbed inputs during the learning process, machine learning gives a defence against targets of adversarial attacks. However, due to the iterative optimisation method, the adversarial training reduces vulnerability to adversarial attacks and stops the leakage of information by the members. Furthermore, Papernot et al [56] noted that adversarial training is useful in protecting from black-box attacks, whereby the opponents attempt to get information from the output of the model without knowing the various parameters of the model. As a result, adversarial training shields machine learning models from such attacks hence preserving the confidentiality of data and improving the privacy of the same.

4. Regulatory and Ethical Considerations

4.1. Ethical Principles in AI and ML

AI and ML systems' design and implementation are subject to several ethical principles that govern their use. These principles include fairness and anti-discrimination [52], transparency and explainability [57], accountability [58], and privacy and data protection [52]. All these ethical principles fundamentally play a crucial role in guiding the leadership in the right deployment and application of AI & ML systems. Therefore, by addressing the fairness, transparency, accountability, and privacy issues, companies can ensure that their AI technologies positively contribute to society and mitigate the potential challenges.

4.2. Impact of Regulations on AI/ML Development

The regulatory frameworks significantly impact the AI/ML technologies development and deployment in multiple ways. These include compliance costs [59], Innovation Constraints [60], enhanced Trust [61],[62], [57] and Global variability [61], [63].

5. Future directions and Emerging trends

5.1. Privacy-Preserving AI/ML Algorithms

According to [64], AI and machine learning domains are progressing rapidly, necessitating the importance of robust privacy-preserving algorithms to take care of data security. However, among these advancements, differential privacy and federated learning are viewed as key technologies in these innovations. Differential privacy is a mathematical approach that can protect the individual entries of a database by adding calibrated noise distortion [65]. This noise is useful to prevent leakage of information of any specific user. In federated learning, the model is trained at several devices or servers that possess small subsets of local data samples [66]. This reduces the multiple copy problem and significantly lowers the threat of getting the data breached since raw data is found on local computers. Federated learning best fits sectors like healthcare and finance, where data sensitivity is important. Thus, it makes it easy for organisations to harness the machine learning benefits, while complying with the legal requirements on privacy rights. Aside from differential privacy and

federated learning, other approaches like homomorphic encryption and secure multi-party computation are also employed in AI/ML with the aim of achieving privacy preservation [67].

5.2. Advances in Cryptographic Techniques

Padmanaban [42] highlighted that the improvements in the cryptographic procedures are crucial for the protection of data processing in AI and ML. Among these innovations, homomorphic encryption and secure multi-party computation (SMPC) are the prominent ones. Homomorphic encryption enables computations to be done straight on the encrypted data without the use of decryption to be made prior to this process [46]. This capability is revolutionary in the aspect of guaranteeing the security of information secret up to the time when the computation is complete. For instance, in medical research, patient records can be encrypted and later used for analysis without revealing the raw data, thereby maintaining patient privacy. Secure multi-party computation, on the other hand, allows multiple parties to collaboratively compute a function on their inputs where the inputs of the other parties remain unknown to them [68]. No information is shared with another party; the information is kept encrypted in each of the parties' databases. This technique is very useful when the various data collected from different sources need to be integrated and analysed while observing data privacy. For instance, the banks can apply SMPC to jointly identify the fraud patterns in their collective database even without sharing their client's details. The continuous advancement of such techniques is driven by the progressive need to enhance data security and privacy in AI and ML applications.

5.3. Decentralised AI/ML Models

As per Wylde et al [69], the shift towards decentralisation of the AI/ML models marks a significant attempt to achieve better data privacy and security. Unlike the traditional centralised model that relies only on one server to solve the problem, in the decentralised models such tasks are solved by several nodes. This structural design eliminates the problem of the data breach and solves the issues that come with single centralisation points. By keeping the data localised on individual nodes, decentralised models reduce the need to transfer sensitive data to the central repository.

Similar to centralised AI/ML models, decentralised models also present protection from cyber-attacks [70]. In centralised systems, a successful invasion of the central server will lead to the corruption of the entire data set and model. However, in a decentralised system, if an attacker gains access to a single node, they cannot access all the information, thereby enhancing the security. In addition, they are more effective and reduce latency by processing data near the source, which is crucial in real-time applications such as autonomous cars and smart cities.

5.4. Policy Developments and Global Cooperation

Future advancement of AI & ML heavily depends on the new policy trends and corresponding cooperation platforms. Thus, with the help of AI & ML tools in constant evolution, the need for robust regulatory measures to ensure data security and privacy to the data appears to be a necessity. Regulatory measures such as the General Data Protection Regulation (GDPR) in the European Union and the California Consumer Privacy Act (CCPA) in the United States have tremendously described data protection laws in terms of legislation and rules [71]. These regulations mandate strict rules on the processing of personal data, like how it is to be collected and stored, thus increasing pressure on organisations to apply firmer modes of privacy.

For instance, GDPR sets out several rigorous measures for acquiring consent for controlling an individual's data, limited data collection, and getting individual data together with erasing it [72]. Similarly, the CCPA also grants California Residents other additional rights concerning their personal information, including the right to know what data is being collected and the right to opt out, among others [72].

Regulation of AI and ML cannot be accomplished separately. Privacy issues are universal today, and different laws in various countries can cause myriad problems for global corporations. Hence, there is a need for international collaboration in terms of synchronising rules and enabling the exchange of data across borders efficiently. As a result, countries working together can result in the formulation of standard approaches and

practices since it reduces the variability of regulation [73].

6. Conclusion

Recent developments in artificial intelligence (AI) and machine learning (ML) have ushered in a new era of business functions, guiding efficiency, innovation, and insights. Although these technologies' use of personal data has been widely embraced, it has led to numerous privacy concerns. The studies have revealed some generic privacy threats unique to AI and ML applications, including Overfitting, data leakage, unauthorised access, re-identification, model inversion attacks, privacy audits and membership inference attacks.

Such challenges and various privacy-preserving solutions have emerged, with their strengths and weaknesses. Differential privacy is one of the most well-known approaches, grounded on mathematical distribution, noise, or information randomness. As for its use, data protection generated by this method refers to personal information and can be used for large populations of people. The other is a secure federated learning approach, a decentralised methodology employed during ML training. This enables local algorithms to be trained using local data without exchanging raw data, which ensures data privacy in the best interests of distributed datasets.

Moreover, homomorphic encryption also introduces another new type of solution since computations on specific information are possible. At the same time, they are encrypted, and therefore, there is no need for the actual data to be decrypted in some physical way [46]. These are fundamental approaches in ensuring privacy in AI and ML as the use of data rises.

REFERENCES

1. Sandeep, S.R., Ahamad, S., Saxena, D., Srivastava, K., Jaiswal, S. and Bora, A., 2022. To understand the relationship between Machine learning and Artificial intelligence in large and diversified business organisations. Materials Today: Proceedings, 56, pp.2082-2086.
2. Sharma, N., Oriaku, E.A. and Oriaku, N., 2020. Cost and effects of data breaches, precautions, and disclosure laws. International Journal of Emerging Trends in Social Sciences, 8(1), pp.33-41.
3. Bandara, R., Fernando, M. and Akter, S., 2020. Privacy concerns in E-commerce: A taxonomy and a future research agenda. Electronic Markets, 30(3), pp.629-647.
4. Ribeiro-Navarrete, S., Saura, J.R. and Palacios-Marqués, D., 2021. Towards a new era of mass data collection: Assessing pandemic surveillance technologies to preserve user privacy. Technological Forecasting and Social Change, 167, p.120681.
5. Martínez-Fernández, S., Bogner, J., Franch, X., Oriol, M., Siebert, J., Trendowicz, A., Vollmer, A.M. and Wagner, S., 2022. Software engineering for AI-based systems: a survey. ACM Transactions on Software Engineering and Methodology (TOSEM), 31(2), pp.1-59.
6. Xu, A., Baysari, M.T., Stocker, S.L., Leow, LJ, Day, RO and Carland, J.E., 2020. Researchers' views on, and experiences with, the requirement to obtain informed consent in research involving human participants: a qualitative study. BMC medical ethics, 21, pp.1-11.
7. Andreotta, A.J., Kirkham, N. and Rizzi, M., 2022. AI, big data, and the future of consent. Ai & Society, 37(4), pp.1715-1728.
8. Zhu, T., Ye, D., Wang, W., Zhou, W. and Philip, S.Y., 2020. More than privacy: Applying differential privacy in key areas of artificial intelligence. IEEE Transactions on Knowledge and Data Engineering, 34(6), pp.2824-2843.
9. Androniceanu, A., 2021. Transparency in public administration as a challenge for a good democratic governance. Revista» Administratie si Management Public «(RAMP), (36), pp.149-164.
10. Kostopoulou, A., 2022. Artificial Intelligence and Personal Data: Topical Issues on the Occasion of the EU AI ACT (Doctoral dissertation, University of Piraeus (Greece)).
11. Mushtaq, MS, Mushtaq, M.Y., Iqbal, M.W. and Hussain, S.A., 2022. Security, integrity, and Privacy of cloud computing and big data. In security and privacy trends in cloud computing and big data (pp. 19-51). CRC Press.

12. Seth, B., Dalal, S., Jaglan, V., Le, D.N., Mohan, S. and Srivastava, G., 2022. Integrating encryption techniques for secure data storage in the cloud. Transactions on Emerging Telecommunications Technologies, 33(4), p.e4108.
13. Al-Matari, O.M., Helal, I.M., Mazen, S.A. and Elhennawy, S., 2021. Integrated framework for cybersecurity auditing. Information Security Journal: A Global Perspective, 30(4), pp.189-204.
14. Thapa, C. and Camtepe, S., 2021. Precision health data: Requirements, challenges and existing techniques for data security and Privacy. Computers in biology and medicine, 129, p.104130.
15. Suleski, T., Ahmed, M., Yang, W. and Wang, E., 2023. A review of multi-factor authentication in the Internet of Healthcare Things. Digital Health, 9, p.20552076231177144.
16. Silva, P., Gonçalves, C., Antunes, N., Curado, M. and Walek, B., 2022. Privacy risk assessment and privacy-preserving data monitoring. Expert Systems with Applications, 200, p.116867.
17. Quach, S., Thaichon, P., Martin, K.D., Weaven, S. and Palmatier, R.W., 2022. Digital technologies: Tensions in privacy and data. Journal of the Academy of Marketing Science, 50(6), pp.1299-1323.
18. Fizur, E., 2020. Long term data retention (Doctoral dissertation, Rutgers University-Camden Graduate School).
19. Mohammad, N., 2021. Data Integrity and Cost Optimization in Cloud Migration. International Journal of Information Technology & Management Information System (IJITMIS), 12, pp.44-56.
20. Zhang, X., Chan, F.T., Yan, C. and Bose, I., 2022. Towards risk-aware artificial intelligence and machine learning systems: An overview. Decision Support Systems, 159, p.113800.
21. Tirumala, K., Markosyan, A., Zettlemoyer, L. and Aghajanyan, A., 2022. Memorization without overfitting: Analyzing the training dynamics of large language models. Advances in Neural Information Processing Systems, 35, pp.38274-38290.
22. Moradi, R., Berangi, R. and Minaei, B., 2020. A survey of regularization strategies for deep models. Artificial Intelligence Review, 53(6), pp.3947-3986.
23. Ye, J., Maddi, A., Murakonda, S.K., Bindschaedler, V. and Shokri, R., 2022, November. Enhanced membership inference attacks against machine learning models. In Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (pp. 3093-3106).
24. Shafee, A. and Awaad, T.A., 2021. Privacy attacks against deep learning models and their countermeasures. Journal of Systems Architecture, 114, p.101940.
25. Hossain, M.T., Afrin, R. and Biswas, M.A.A., 2024. A Review on Attacks against Artificial Intelligence (AI) and Their Defence Image Recognition and Generation Machine Learning, Artificial Intelligence. Control Systems and Optimization Letters, 2(1), pp.52-59.
26. Ivanovs, M., Kadikis, R. and Ozols, K., 2021. Perturbation-based methods for explaining deep neural networks: A survey. Pattern Recognition Letters, 150, pp.228-234.
27. Liu, B., Ding, M., Shaham, S., Rahayu, W., Farokhi, F. and Lin, Z., 2021. When machine learning meets Privacy: A survey and outlook. ACM Computing Surveys (CSUR), 54(2), pp.1-36.
28. Di Minin, E., Fink, C., Hausmann, A., Kremer, J. and Kulkarni, R., 2021. How to address data privacy concerns when using social media data in conservation science. Conservation Biology, 35(2), pp.437-446.
29. Martinez, D. and Herrera, S., 2023. Examining the Ethical and Legal Challenges of Anonymized Data Sharing in the Era of Big Data Analytics. Journal of Sustainable Technologies and Infrastructure Planning, 7(5), pp.59-77.
30. Ni, C., Cang, L.S., Gope, P. and Min, G., 2022. Data anonymization evaluation for big data and IoT environment. Information Sciences, 605, pp.381-392.
31. Gowda, VT, 2023. Methods to achieve t-closeness for Privacy preserving data publishing (Doctoral dissertation, Wichita State University).
32. Aslan, Ö., Aktuğ, S.S., Ozkan-Okay, M., Yilmaz, A.A. and Akin, E., 2023. A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions. Electronics, 12(6), p.1333
33. Balisane, H., Egho-Promise, E. I., Lyada, E., & Aina, F. (2024). Towards Improved Threat Mitigation In Digital Environments: A Comprehensive Framework For Cybersecurity Enhancement. International Journal of Research -GRANTHAALAYAH, 12(5), 108–123
34. Ferra, F., Wagner, I., Boiten, E., Hadlington, L., Psychoula, I. and Snape, R., 2020. Challenges in assessing privacy impact: Tales from the front lines. Security and Privacy, 3(2),

35. Raab, C.D., 2020. Information privacy, impact assessment, and the place of ethics. Computer Law & Security Review, 37, p.105404.
36. La Torre, M., Botes, V.L., Dumay, J. and Odendaal, E., 2021. Protecting a new Achilles heel: the role of auditors within the practice of data protection. Managerial Auditing Journal, 36(2), pp.218-239.
37. Murakami, T. and Takahashi, K., 2020. Toward evaluating re-identification risks in the local privacy model. arXiv preprint arXiv:2010.08238.
38. Majeed, A., Khan, S. and Hwang, S.O., 2022. Toward privacy preservation using clustering based anonymization: recent advances and future research outlook. IEEE Access, 10, pp.53066-53097.
39. Templ, M., Kanjala, C. and Siems, I., 2022. Privacy of study participants in open-access health and demographi Schaller c surveillance system data: Requirements analysis for data anonymization. JMIR Public Health and Surveillance, 8(9), p.e34472.
40. Zuo, Z., Watson, M., Budgen, D., Hall, R., Kennelly, C. and Al Moubayed, N., 2021. Data anonymization for pervasive health care: systematic literature mapping study. JMIR medical informatics, 9(10), p.e29871.
41. Bowen, C.M. and Garfinkel, S., 2021. Philosophy of differential Privacy. Notices of the American Mathematical Society, 68(10), pp.1727-39.
42. Padmanaban, H., 2024. Privacy-Preserving Architectures for AI/ML Applications: Methods, Balances, and Illustrations. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 3(1), pp.235-245.
43. Mammen, P. M. (2021). Federated learning: Opportunities and challenges. arXiv preprint arXiv:2101.05428.
44. Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. Future Generation Computer Systems, 115, 619-640.
45. Li, H., Ge, L., & Tian, L. (2024). Survey: federated learning data security and privacy-preserving in edge-Internet of Things. Artificial Intelligence Review, 57(5). https://doi.org/10.1007/s10462-024-10774-7
46. Asante, G., Hayfron-Acquah, J. B., Asante, M., & Dagadu, J. C. (2022). A symmetric, probabilistic, non-circuit based fully homomorphic encryption scheme. International Journal of Computer Networks and Applications (IJCNA), 9(2), 160-168.
47. Wang, Y., Liang, X., Hei, X., Ji, W., & Zhu, L. (2021). Deep learning data privacy protection based on homomorphic encryption in AIoT. Mobile Information Systems, 2021(1), 5510857.
48. Rahman, M. S., Khalil, I., Atiquzzaman, M., & Yi, X. (2020). Towards privacy preserving AI based composition framework in edge networks using fully homomorphic encryption. Engineering Applications of Artificial Intelligence, 94, 103737.
49. Yaji, S., Bangera, K., & Neelima, B. (2018, December). Privacy preserving in blockchain based on partial homomorphic encryption system for AI applications. In 2018 IEEE 25th International Conference on High Performance Computing Workshops (HiPCW) (pp. 81-85). IEEE.
50. Skarkala, M. E., Maragoudakis, M., Gritzalis, S., & Mitrou, L. (2021). PPDM-TAN: A Privacy-Preserving Multi-Party Classifier. Computation 2021, 9, 6.
51. Schaller, F. (2022). Adversarial robustness in computer vision: A review. IEEE Transactions on Neural Networks and Learning Systems, 33(1), 201–215. doi: 10.1109/TNNLS.2021.3089451
52. Zhou, J., Chen, F., Berry, A., Reed, M., Zhang, S. and Savage, S., 2020, December. A survey on ethical principles of AI and implementations. In 2020 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 3010-3017). IEEE.
53. Pilton, A., Riello, A., & Veronese, L. (2021). Towards robust neural networks with adversarial training. Neural Computing and Applications, 33(10), 5327–5337. doi: 10.1007/s00521-021-05514-9
54. Goodfellow, I., Shlens, J., & Szegedy, C. (2014, November). Explaining and harnessing adversarial examples. In Proceedings of the 31st International Conference on Machine Learning (Vol. 32, pp. 2979-2987).
55. Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. Proceedings of the 35th International Conference on Machine Learning, 80, 2925–2934.
56. Papernot, N., Mcdaniel, P., Goodfellow, I.J., Jha, S., Celik, Z.B., & Swami, A. (2016). Practical Black-

Box Attacks against Machine Learning. Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security.

57. Mittelstadt, B., 2019. Principles alone cannot guarantee ethical AI. Nature machine intelligence, 1(11), pp.501-507.
58. Currie, G., Hawk, K.E. and Rohren, E.M., 2020. Ethical principles for the application of artificial intelligence (AI) in nuclear medicine. European Journal of Nuclear Medicine and Molecular Imaging, 47, pp.748-752.
59. Padmanaban, H., 2024. Navigating the Complexity of Regulations: Harnessing AI/ML for Precise Reporting. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 3(1), pp.49-61
60. Padmanaban, H., 2023. Navigating the intricacies of regulations: Leveraging AI/ML for Accurate Reporting. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(3), pp.401-412.
61. Padmanaban, H., 2024. Revolutionizing Regulatory Reporting through AI/ML: Approaches for Enhanced Compliance and Efficiency. Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023, 2(1), pp.71-90.
62. Cobbe, J., Lee, M.S.A. and Singh, J., 2021, March. Reviewable automated decision-making: A framework for accountable algorithmic systems. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 598-609).
63. Abbott, K.W. and Snidal, D., 2021. Strengthening international regulation through transnational new governance: Overcoming the orchestration deficit. In The spectrum of international institutions (pp. 95-139). Routledge.
64. Kaissis, G.A., Makowski, M.R., Rückert, D. and Braren, R.F., 2020. Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 2(6), pp.305-311.
65. Xiong, X., Liu, S., Li, D., Cai, Z. and Niu, X., 2020. A comprehensive survey on local differential privacy. Security and Communication Networks, 2020(1), p.8829523.
66. Imteaj, A., Thakker, U., Wang, S., Li, J. and Amini, M.H., 2021. A survey on federated learning for resource-constrained IoT devices. IEEE Internet of Things Journal, 9(1), pp.1-24.
67. Truong, N., Sun, K., Wang, S., Guitton, F. and Guo, Y., 2021. Privacy preservation in federated learning: An insightful survey from the GDPR perspective. Computers & Security, 110, p.102402.
68. Feng, D. and Yang, K., 2022. Concretely efficient secure multi-party computation protocols: survey and more. Security and Safety, 1, p.2021001.
69. Wylde, V., Rawindaran, N., Lawrence, J., Balasubramanian, R., Prakash, E., Jayal, A., Khan, I., Hewage, C. and Platts, J., 2022. Cybersecurity, data privacy and blockchain: A review. SN computer science, 3(2), p.127.
70. Singh, J., Wazid, M., Das, A.K., Chamola, V. and Guizani, M., 2022. Machine learning security attacks and defense approaches for emerging cyber physical applications: A comprehensive survey. Computer Communications, 192, pp.316-331.
71. Choi, K.H.M., 2020. A Critical Juncture in Data Protection Standards: Comparing Data Protection Legislation in the United States and the European Union (Doctoral dissertation, 서울대학교 대학원).
72. Saquella, A.J., 2020. PERSONAL DATA VULNERABILITY. Jurimetrics, 60(2), pp.215-245.
73. Mone, V. and Mitharwal, S., 2024. Guardians of privacy: exploring the viability of a United Nations-backed global data governance. International Journal of Intellectual Property Management, 14(2), pp.194-216.
74. Jagarlamudi, G. K., Yazdinejad, A., Parizi, R. M., & Pouriyeh, S. (2024). Exploring privacy measurement in federated learning. The Journal of Supercomputing, 80(8), 10511-10551.

Authors

Ehigiator Egho-Promise is a Dean for West Africa at European-American University; SPO(Course Lead) and Lead Internal Verifier at City of Oxford College and University Centre, Visiting Computer Science Lecturer at Birmingham Newman University, external examiner for graduate and postgraduate Cybersecurity students at National Open University of Nigeria; an academic reviewer and editorial board member of some reputable international journals; professional member of British Computer Society, Association for Computing Machinery, Information Systems Audit and Control Association, Chartered Fellow at Chartered Institute of Strategic Managers and Leaders; International Eminent Peace Ambassador, Global SDGs Advocate Special Chartered Membership Award UN SDGs.

He has over 20 years' industry experience which cut across banking, IT, Telecommunication, oil & Gas sectors. Furthermore, he has more than 10 years teaching and research experience and has authored and coauthored over 30 articles in peer reviewed journals.

He holds several qualifications which include but not limited to the following: PhD in Data Communication and Networking, PhD in Business Administration, Master of Science in Information Technology, Master Business Administration, Bachelor of Science in Computer Science, Higher Diploma in Accounting, Higher Diploma in Electrical/Electronic Engineering Technology, Diploma in Data Processing, and Ordinary Diploma in Financial Studies.

His research areas include Cybersecurity, Cyber Forensics, Telecommunication Network, Data Communication and Networking, Artificial Intelligence and Machine Learning

Dr. George Asante is a Senior Lecturer in the Department of Information Technology Education of the Akenten Appiah-Menka University of Skills Training and Entrepreneurial Development, Kumasi, Ghana. He holds a Ph.D. in Computer Science, MPhil in Information Technology and B.Ed in Information Technology. His research areas include Information Security, Digital Forensics, Algorithms, and ICT Education.

Dr Hewa Balisane, Senior Lecturer in Cyber Security at the University of Law. With a distinguished career spanning over 15 years, He completed his Doctoral Thesis at Manchester Metropolitan University where his research focused on information. Dr Balisane has held pivotal roles such as Head of Departments, Assistant Dean, Dean of Faculty, and Vice Rector at different universities in the Middle East. His extensive experience in the Higher Education sector is marked by an international profile and a commitment to delivering high-quality academic teaching, research, and leadership both nationally and internationally. His research interests encompass a broad range of subjects within the technological domain, including Cyber Security, Biometrics, advanced information technology, multimedia technology, data telecommunications and networks. His dedication to these areas demonstrates his involvement with the forefront of technological advancements that are shaping our modern society. Currently, Dr Balisane conducts research and oversees modules such as ‘Cyber Security for Business’ and ‘Cyber Security Management and Compliance’ at the University of Law Business School. His role as a research lead underscores his proficiency in navigating the intricacies of cyber security within the framework of business operations and compliance.

Folayo. A. Aina is the course leader for the BSc Networks and Security program at the University of Central Lancashire (UCLAN). She received her bachelor's degree in computer science from the University of Ilorin, Nigeria, in 2009, followed by a master's degree in network systems from the University of Sunderland, UK, in 2011. She completed her Ph.D. in Computing at Anglia Ruskin University, Chelmsford, UK, in 2020. Currently, she serves as a Lecturer at the School of Engineering and Computing at UCLAN in Preston, United Kingdom. Her research interests encompass Computer Networks, Cyber Security, Artificial Intelligence, Machine Learning, Wireless Networks, MANET (Mobile Ad hoc Networks), Mobile Communication, and Network Security.

Halima Ibrahim Kure is a Senior Lecturer in Cybersecurity in the School of Engineering and Computing at the University of East London. She previously served as a Lecturer in Cybersecurity at the University of Central Lancashire. Dr. Kure obtained her Ph.D. in Cybersecurity and Risk Management from the University of East London and holds a B.Sc. in Software Engineering and an M.Sc. in Information Security and Computer Forensics from the same institution.

Dr. Kure's areas of expertise encompass a wide range of topics, including Artificial Intelligence-enabled cybersecurity, cyber resilience, cyber threat intelligence, security and risk management, cyber-physical systems, threat modelling, attack vectors, and cloud security management. She possesses extensive knowledge of prominent cybersecurity frameworks and standards, such as ISO 27001, NIST's CIS CSC, and NIST CSF. In her academic roles, Dr. Kure actively engages in research and development initiatives focused on enhancing the security, privacy, and resilience of systems and applications across diverse domains. Her research extends to critical information infrastructure, cyber-physical systems, mobile applications, cloud computing, and medical and healthcare systems. She is committed to advancing the field of cybersecurity through her innovative research, impactful teaching, and dedication to student success.