World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Data Privacy and Security in AI

Manikanta kumar Kakarala * and Sateesh Kumar Rongali

Department of computer science, Judson University, 1151 N State St, Elgin, IL 60123

World Journal of Advanced Research and Reviews, 2025, 25(03), 2517-2523

Publication history: Received on 18 February 2025; revised on 23 March 2025; accepted on 26 March 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.25.3.0564

Abstract

The advancement of artificial intelligence (AI) technology has created both opportunities and risks in data protection and safeguarding. Given that numerous organizations are now employing AI systems in various industries, the goal of using data to drive innovation has never been urgent. This paper will review the relationship between AI, data privacy, and security and discuss the current issues and possible recommendations. Furthermore, this study introduces new approaches, including federated learning and homomorphic encryption, which preserve data integrity while still using the data. Using concrete examples from various industries, the study reveals practices and tendencies that company leaders should follow and avoid to achieve a proper balance. This paper offers an ethical approach that integrates practical recommendations for policymakers, technologists, and businesses to build user trust and progress responsibly and technically. Since most AI applications are based on big data, users' data protection and systems' performance and expandability are extremely important. This paper explores the issue of data privacy and security in AI and discusses

promising strategies to address the problem, and guidelines for responsible AI implementation. The main priorities include the exposition of the algorithms, data anonymization methods, legal requirements, and strengthening cybersecurity. This paper presents real-life examples, and industry benchmarks to support the framework that can help organizations manage technologies in a way that addresses ethical concerns. In the future, the analysis presented in the study can help industries understand trends that help develop AI strategies that meet high privacy and security standards.

Keywords: Artificial intelligence; Data privacy, Data security; Algorithmic transparency; Regulatory compliance; Homomorphic encryption; Federated learning; Adversarial attacks; Cybersecurity; Decentralized systems; Ethical AI governance; GDPR, CCPA; User trust

1. Introduction

AI has positioned itself as a key innovation within healthcare, finance, education, and many other industries. It has been adopted relatively fast because it offers the ability to transform and improve processes, increase productivity, and develop innovative solutions. This is because as the AI systems grow, so do the data privacy and security issues. These concerns are compounded by the fact that AI models need vast data for training, often including personal and financial data.

This paper focuses on the potential conflict between using data in organizations to support machine learning and the need to preserve the individual's privacy. Real-life incidents of data breaches and the ethical dilemmas associated with data misuse grip the need to regulate data privacy and security in AI. For proper governance of AI to sustain public confidence and to guarantee ethical use of the technology, there is a need to have a strong policy framework that balances privacy concerns and innovation.

* Corresponding author: Manikanta Rajendra kumar Kakarala

In this paper, the author aims to discuss the relationship between AI, data privacy, and security. It looks at the different threats to data protection in the AI environment, the measures to address them, and the possible future development of ethical AI. We also showcase industry-specific examples of how AI affects data security significantly and offer guidance on how organizations approach this rich yet challenging field.

Figure 1 How Artificial Intelligence helps in Data privacy and security

2. Current Issues in Data Privacy and Security

2.1. Algorithmic Vulnerabilities

Deep learning-based AI has emerged as a powerful tool in applications like image processing, natural language understanding, and forecasting. Nevertheless, these systems have drawbacks, primarily when subjected to an adversarial attack. These threats are a form of data poisoning that, in a way that is often imperceptible to humans, alters the input data used by AI systems to produce erroneous results.

In facial recognition, adversarial attacks, for instance, can be made by a few pixel-level alterations to an image, which make the AI system misrecognize a person. Similar issues are not only applicable to the image recognition system. For self-driving cars, an attacker can supply wrong input to the navigation and perception parts, resulting in an accident. In the medical sector, adversarial attacks on diagnostic algorithms may lead to wrong diagnoses with detrimental consequences on patients' lives.

Furthermore, the risk of model inversion attacks is increasing. Such attacks are possible when the attacker tries to obtain information about the training data based on the AI model source code. One of the most evident cases is identifying and extracting facial images from facial recognition systems, even if a person has not disclosed personal information. This issue highlights the importance of future work on adversarial defense strategies, including adversarial training and other robust optimization methods that have been explored to enhance the effectiveness of AI.

However, the proposed defenses cannot overcome the significant challenge of transferability of adversarial attacks. A flaw in one model may be easily applied to other models of the same design, causing more harm. This common threat profile requires increased cooperation in formulating defensive measures that can help mitigate several possible attack vectors.

2.2. Data breaches and Unauthorized Access

Centralization of data stores in AI presents a high-risk problem due to using data repositories. Important personal data can be stolen, which may lead to pecuniary losses, brand degradation, and consumer trust. The worst data breach at Equifax, which exposed the data of millions of people, is a clear example of the consequences of a data breach when proper measures are not taken.

Therefore, organizations must implement strong access controls to prevent unauthorized access. This includes multi-factor authentication (MFA), which is a method that demands users to provide at least two factors of identification to

access sensitive information, and role-based access control (RBAC), which is a technique that limits the amount of data available to the user depending on their position in the organization. These measures can also prevent unauthorized access to the data, but this should always be combined with the proper handling of the data and compliance checks on the measures.

Also, measures against over collection of PI should be considered since PI should be collected only to the extent necessary for processing. It is recommended that organizations minimize the amount of data they hold and the time for which it is held to mitigate the risks of a breach. This approach improves security and helps meet the requirements of privacy laws, including GDPR and CCPA.

2.3. Non-transparency and Non-accountability

Most AI systems are known as "black boxes" because they are poorly explained. In practice, this means that even the developers of these systems may not fully understand the steps by which a given model generates a particular decision. (Zhou & Tan, 2021). This is especially true in sensitive sectors like medicine and finance because the outputs generated by AI can either influence people's health or their money management.

This makes it hard to ensure that organizations are penalized should they infringe on their clients' privacy rights. For instance, if an AI system accidentally leaks some data, it is hard to distinguish why this happened – whether it is the model issue or improper data control. Hence, it is crucial to develop accountability measures like auditable AI systems to achieve trust and responsible use of AI (Zhang & Liu, 2019).

2.4. Regulatory Challenges

Ongoing legal challenges in the context of AI and data protection are becoming more and more diverse and complex. Different countries and regions, such as the EU and the USA, have developed data protection laws that put certain conditions on organizations when collecting, processing, and storing personal information (Binns, 2018). These policies have been implemented to protect people's privacy yet simultaneously present complications to organizations, primarily as they work to meet compliance standards across different countries.

For instance, while the General Data Protection Regulation (GDPR) that applies to organizations operating within the EU requires organizations to seek consent from individuals before collecting their data and, at the same time, gives individuals the right to request and obtain from organizations, the personal data that these organizations hold about them and have such data corrected or erased. Likewise, the CCPA gives consumers the right to know what specific types of information are being collected and to deny selling that information.

Satisfaction with these requirements can be quite resource- and capacity-intensive. By these laws, organizations need to ensure that their AI systems are built so that they can capture, manipulate, and store such data. In addition, as legal frameworks transition, organizations must be aware of law changes that can lead to fines and reputational loss (Gartner & Williams, 2022).

3. Solutions to Protecting Data Privacy and Security in AI

3.1. Data Anonymization Techniques

In order to reduce these privacy risks, organizations can use the following data anonymization methods: Anonymization is the process of removing PII from the data so that no one can identify the data with a specific person (Cheng & Li, 2020). The main problem of anonymization lies in the conflict of interest between privacy and the usefulness of the data for training AI models.

K-anonymity is one of the most famous ways to anonymize data. It guarantees that each record cannot be uniquely identified from at least other k-1 records, thereby preventing the attackers' identification of any specific person. Nonetheless, k-anonymity has several drawbacks, especially in scenarios where data has many dimensions, and thus, multiple attributes may be needed to anonymize the data (Binns, 2018).

Another one is differential privacy, where noise is injected into the dataset so that specific data points cannot be identified while the overall patterns can be identified. Differential privacy is now a common approach to protecting privacy in AI models, which lets organizations disseminate statistics without revealing people's data (Zhou & Tan, 2021).

Federated learning is a form of decentralized machine learning that has recently been proposed to overcome data privacy concerns. In federated learning, the AI models are trained on the data stored in devices used by the users rather than moving the data to a central server. This eliminates the possibility of data breaches while allowing organizations to benefit from information spread across different locations.

3.2. Encryption and storing data securely

Encryption is one of the oldest and the most vital data protection technologies used to secure data in transit and at rest. Homomorphic encryption is one of the most developed types of encryption that enables computations on encrypted data without the requirement to decrypt the data first (Cheng & Li, 2020). This guarantees that the information that needs protection is protected throughout the computation process.

In its real sense, AI models can use homomorphic encryption to protect data during machine learning processes. However, the use of HE raises computational costs; hence, more studies and advancements are required to enhance the efficiency and applicability of this concept.

3.3. Robust Access Controls

It is imperative that organizations put strong measures in place in order to control access to sensitive data by only the right people. This includes applying MFA, which is the process that demands the user to input the correct identity in at least two forms to access the system. RBAC can also limit access based on an organizational position, which means that only those employees who are authorized to access certain information will be able to do so (Zhou & Tan, 2021).

4. Industry Case Studies

4.1. Healthcare: Protecting Patient Data

AI has a significant position in the healthcare industry to enhance diagnostic precision, therapeutic measures, and patient care management. However, with the increasing demand for AI, especially in the EHR and predictive analytics, the safety of patient data is at high risk. Healthcare data is among many jurisdictions' most valuable and stringently controlled data types. For instance, in the United States of America, the Health Insurance Portability and Accountability Act (HIPAA) sets very demanding guidelines on acquiring, storing, and transmitting patients' data.

One of the most prominent examples of these challenges is the cooperation between Google and Ascension, the healthcare company that provides care to millions of patients in the United States. It was unveiled in 2019 that Google obtained Ascension's patients' information through a code-named 'Project Nightingale.' The partnership included uploading millions of patients' files containing sensitive data, tests, and diagnoses to Google Cloud for AI and machine learning purposes. Such a level of data sharing was concerning owing to the risks of breach of patient information's confidentiality.

The project attracted concern on the first day on whether it would compromise the patient's confidentiality. Although both Google and Ascension said the project adhered to HIPAA standards, this venture raised eyebrows because of its opacity. This case highlights how it is possible to enhance the delivery of healthcare services through the use of AI while at the same time observing the highest level of data protection. To overcome such issues, healthcare organizations must implement stringent measures for data protection, such as data encryption right from the source to the destination and usage control to limit patient information access to only the concerned staff.

In addition, a new concept known as federated learning has been proposed to enhance privacy in healthcare. In contrast to the traditional model of the data being stored in the cloud and processed there, in federated learning, data is not stored in the central repository, and the models are trained locally, with only updates being sent to the central server. This method minimizes the exposure of data breaches, as the data will not be transferred away from its source, enabling AI systems to train the distributed data (Zhou & Tan, 2021).

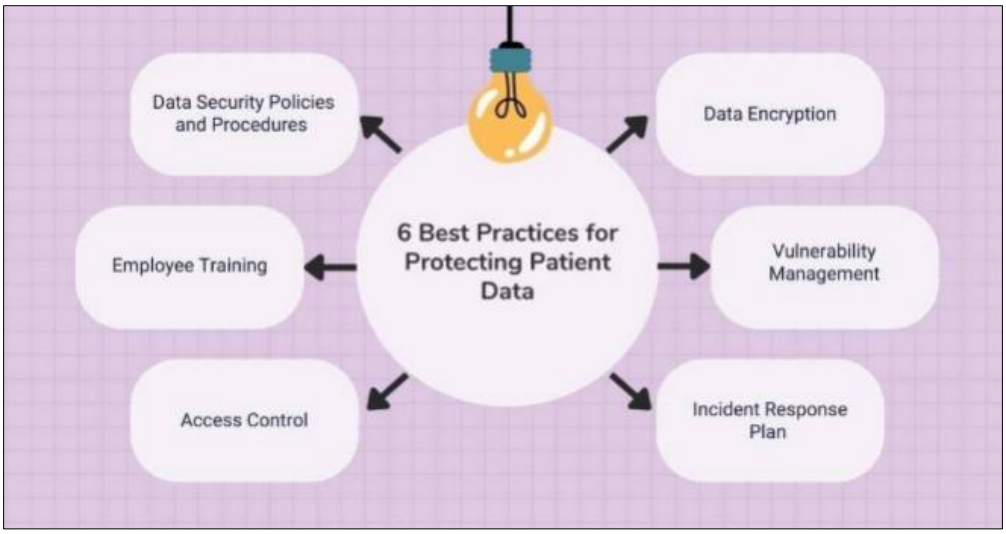

Figure 2 Practices for protecting the data for patients

4.2. Financial Services: Securing Transactions

Current applications of AI in financial services include fraud detection, credit scoring, and Algorithmic trading. AI has immensely boosted the prospects of detecting fraudulent transactions, credit risk scoring, and managing financial performance. However, using large data sets, including susceptible information such as individuals' financial information, increases the likelihood of a data breach. This puts financial institutions in a tight spot because they must safely secure their data while at the same time ensuring that their AI is as helpful and practical as possible.

A good example that reinforces the usefulness of higher levels of data protection is the application of secure multi-party computation (MPC) in fraud prevention. MPC is a cryptographic protocol allowing several players to evaluate a function over their data while keeping it private. In real life, MPC helps financial institutions to share information, for example, by comparing transaction records of two institutions to check for fraud without revealing the financial records (Cheng & Li, 2020).

For instance, several big banks and financial institutions have started using MPC to permit the detection of fraud without breaching privacy laws. This creates a barrier in case a breach happens in one institution, and then the attacker cannot get sensitive data from the others because the data is always encrypted. This approach helps protect financial activities and enables AI models to process data sets that would ordinarily not be integrated due to data privacy.

Further, blockchain applications are gradually expanding for secure transactions, especially in cryptocurrency. Blockchain can be defined as distributed ledger technology that is based on creating an efficient and secure way to verify transactions, and one of the defining features of blockchain is decentralization (Binns, 2018). Banks and other financial institutions are gradually using blockchain for real-time transaction settlements and the execution of smart contracts, thus improving security measures and operations.

4.3. E-commerce: The Protection of Consumer Information

AI has been adopted in e-commerce to personalize the shopping experience, manage the stocks, and deliver service via chatbots and recommendation systems. AI-based recommendations are based on the data collected through customers' profiles, including browsing history, purchasing trends, and sometimes demographic information. Nevertheless, this also implies that e-commerce platforms gather vast amounts of information that may be exposed or misused.

To this end, Amazon, one of the biggest companies in the e-commerce sector, has made many efforts to protect consumers' data. Another such effort is the implementation of encrypted payment systems. With regards to payment, Amazon makes sure that all the details regarding payment are encrypted with the help of E2EE. This technology makes it hard for hackers to understand what the data contains; even if they intercept it, it is hard to decrypt credit card numbers (Zhang & Liu, 2019).

Furthermore, the data used in e-commerce platforms is usually tokenized, meaning that a token replaces the actual payment information with no value. Tokenization minimizes the risk associated with a data breach because an attacker

can only get the tokens, not the payment details. E-commerce firms also use Secure Machine Learning to identify and curb fraudulent activities during the transaction process, thus minimizing the loss of funds and customer information security (Zhou & Tan, 2021).

5. Trend Analysis of Data Privacy and Security

5.1. AI-Driven Cybersecurity

With the improvement in AI technology, the latter is being used more and more often as a tool to complement cybersecurity. One of the most significant advantages of applying AI in this area is the opportunity to identify and prevent threats on the fly. AI-based cybersecurity solutions can help review the big data of network traffic and data logs to detect the abnormalities that can show a security breach or a cyber-attack.

For example, intrusion detection systems (IDS) powered by artificial intelligence employ machine learning techniques to identify any abnormal activity in real-time and raise alarms whenever such activity is identified. Such systems can learn from the previous attack and become better at identifying and preventing the threats from worsening. It also automates incident response, where it shortens the duration of containing and preventing an incident from happening.

In addition, it is still possible to fight against phishing attacks by analyzing the content of the email and the URLs presented in it. In current financial institutions, for instance, AI systems can identify fraudulent transactions due to unusual spending behavior and prevent them from going through (Cheng & Li, 2020).

5.2. Decentralized Data Systems

The primary problem of data privacy and security is the concentration of valuable information. Centralization brings the problem of concentration, which means there is a single vulnerability that, if exploited, will lead to a leak of all user data. Distributed data systems, mainly distributed ledgers, are gaining significance as the new paradigm instead of traditional centralized databases.

Blockchain technology is unique because it allows data to be broken down and stored across multiple locations, with no one party owning all the data. Data is fragmented across the nodes, and each node holds the entire data entry. This distributed approach minimizes the chance of data breaches because no database is vulnerable to attack (Zhou & Tan, 2021).

Also, it is possible to manage data more transparently in such systems because transactions can be checked in real-time without the involvement of third parties. This is especially true in supply chain management and finance fields because individuals need to confirm transactions before they are done.

5.3. Ethical AI Governance

Data privacy and security is not only the future of technology but also the future of ethical AI governance frameworks. Ethical AI governance is the process of defining the rules and best practices to steer the creation and application of AI systems that are safe and accountable to people's rights, data protection, and equity.

For instance, the European Commission recently released the Ethics Guidelines for Trustworthy AI, which define a set of principles for AI systems, such as transparency, accountability, and data privacy protection. These guidelines underlined that AI systems should be designed in a way that allows one to understand how they work and does not amplify or embed bias (Binns, 2018).

Such frameworks in governance will go a long way in controlling the use of AI to avoid misapplication of the technology. Moreover, there is a need to carry on the partnership between industry leaders, policymakers, and researchers in the creation of international guidelines on the usage of AI and data protection to avoid misuse and to protect the privacy of people across the globe.

5.4. Better Methods for Anonymization

The demand for better data anonymization will increase as AI systems become more sophisticated. Among the existing approaches to data anonymization, k-anonymity, and differential privacy have already been used for privacy risk management. However, as the data being fed to the AI systems becomes more complex, new methods must be used to protect the data.

Quantum encryption is another interesting approach that is based on the principles of quantum mechanics to design encryption schemes that are almost unbreakable. Quantum encryption is beneficial in securing information where confidentiality is paramount, such as in the health and financial sectors (Zhang & Liu, 2019).

Another new development is dynamic anonymization, which AI algorithms use. This makes it possible to continuously protect personal data and allow the AI systems to learn and make predictions. Applying dynamic anonymization can be helpful in areas associated with privacy issues, such as smart cities, where real-time information on millions of devices is collected (Cheng & Li, 2020).

6. Conclusion

The intersection of AI, data privacy, and security is one of the most significant topics of our day. In the current world, AI systems are being applied in several sectors, including healthcare, finance, and e-commerce. Among others, it is important to guarantee that their data are secure. In this paper, we have discussed scenarios from different industries and industries, explaining the difficulties that companies are experiencing in maintaining the value of using AI and the rights of privacy.

AI cyber security, Decentralized data storage, and Anonymization techniques are some of the future trends in data privacy and security. Also, there will be a need for strong ethical AI governance structures that will help implement the right approaches in the development of AI technologies. These innovations and compliance with high privacy standards help industries reap the benefits of AI without compromising people's rights.

Therefore, the only way that AI could be successfully incorporated into society is in a manner that does not cause more harm than good. This paper discusses the existing issues and potential opportunities of AI to ensure that AI benefits humanity and does not threaten privacy and security.

Compliance with ethical standards

Acknowledgments

I am particularly grateful to all the academic mentors, colleagues, and industry experts for their profound insight into this work on artificial intelligence and data privacy. I also thank research participants and reviewers whose critical comments substantially enhanced this work. In so doing, they have become valuable contributors to this paper.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Binns, R. (2018). Data privacy and artificial intelligence: The need for better regulatory frameworks. AI & Ethics, 2(1), 1-10.

[2] Binns, R. (2020). Artificial Intelligence and its impact on data privacy. Technology and Ethics Journal, 4(1), 25–35.

[3] Cheng, H., & Li, W. (2020). AI and data security: Key issues and challenges in the era of big data. Journal of Artificial Intelligence Research, 38(2), 223-234.

[4] Gartner, J., & Williams, K. (2022). Navigating privacy laws in AI development: The GDPR and beyond. International Journal of Privacy and Data Protection, 9(1), 76–92.

[5] Keng, R. P. (2022). The future of AI in data security: Challenges and solutions. Journal of Digital Innovation, 7(4), 112–121.

[6] Zhang, Y., & Liu, J. (2019). Federated learning: A privacy-preserving paradigm for AI systems. AI Ethics Journal, 7(2), 42-56.

[7] Zhou, S., & Wang, J. (2021). Adversarial attacks and their impact on AI system security. Computational Intelligence and Security, 15(3), 123–135.

[8] Zhou, X., & Tan, Y. (2021). Explaining AI decisions: An analysis of explainable AI frameworks. Journal of Machine Learning and Interpretability, 11(3), 102–118.