Artificial intelligence (AI) and machine learning (ML) in data protection.

Hanish Yarasani Department of MSIA Saint Cloud State University St. Cloud, Minnesota, 56301 hanish.yarasani@stcloudstate.edu	Anushka U. Hewarathna Department of MSIA Saint Cloud State University St. Cloud, Minnesota, 56301 auhewarathna@go.stcloudstate.edu	Priyanka Akula Department of MSIA Saint Cloud State University St. Cloud, Minnesota, 56301 Priyanka.akula@go.stcloudstate.edu

Abstract

Data protection is one of the most important challenges as it involves obtaining, storing, analyzing, and utilizing a vast amount of data that contains sensitive and personally identifiable information that must be safeguarded. Artificial intelligence (AI) and machine learning (ML) technologies have developed into powerful tools for enhancing data protection by enabling more accurate and timely analysis, monitoring, and reaction to security issues. However, using AI and ML for data protection also has drawbacks, including ethical issues, algorithmic prejudice, and privacy issues. Algorithmic bias may lead to discriminatory outcomes, and ethical issues include the fairness and openness of the algorithms that are utilized. Data collection and analysis without individuals' express consent raises privacy issues. Therefore, there is a need for greater accountability and transparency in how AI and ML are used to secure data.This review article finishes by outlining prospective directions for further study, including the creation of more moral and open AI and ML algorithms and the investigation of novel uses for these tools in data security.This paper emphasizes the benefits and drawbacks of these technologies as well as their potential to improve the field by providing a complete analysis of current trends and future possibilities in the application of AI and ML in data protection.

Keywords: Artificial Intelligence, Machine Learning

Introduction

Data protection using artificial intelligence (AI) and machine learning (ML) refers to the application of these technologies to improve the security and protection of data. Data protection has been elevated to a top priority for both organizations and people due to the growth in data generation and collection. The way data is safeguarded has the potential to change, becoming more effective and secure thanks to AI and ML. Data protection applications that leverage AI and ML include fraud detection, identifying security breaches, and encrypting sensitive information. AI and ML are able to encrypt sensitive data using specific encryption algorithms and keys by analyzing big datasets to find patterns that may point to fraudulent actions, monitor network activity to find security breaches and discover trends that may point to further malicious activities.

The necessity of data security in the current digital era is what inspired our research on artificial intelligence and machine learning in data protection. Data breaches and unauthorized access are becoming increasingly risky due to the exponential expansion in data collecting and utilization. To solve these issues and improve data security, the application of AI and ML in data protection offers a viable option. This study attempts to evaluate previous research on the uses of AI and ML in data protection as well as the difficulties involved in their implementation.

The trouble with incorporating artificial intelligence (AI) and machine learning (ML) in data protection is that it is becoming more and more important. While using AI and ML can improve data security, there are drawbacks such ethical issues, algorithmic prejudice, and privacy worries. In the current digital era, data protection is a crucial concern because of the growing quantity of data being created and gathered. The application of AI and ML is a viable way to address these difficulties because the dangers related to data breaches and illegal access are high. To make sure that these technologies are used responsibly and ethically, privacy issues and ethical issues must be addressed.

The use of AI and ML in data protection, as well as the difficulties involved, have been examined in a number of research. Comprehensive study is still required to fully grasp the possible advantages and difficulties of using AI and ML for data protection, as well as to source any prospective areas for development.

The present research on the uses of AI and ML for data protection, as well as the difficulties in implementing them, is reviewed in this study. It points up possible topics for further study and development, including the creation of more moral and open algorithms and the investigation of novel uses for these technologies in the field of data security.

Despite the fact that there have been several studies on the use of AI and ML in data protection, this paper contributes by doing a thorough evaluation of the literature and identifying promising areas for further study and development. It summarizes the status of the field's knowledge at the moment and offers a direction for further study.

Since it covers important issues related to the application of AI and ML in data protection, the contribution is not trivial. Researchers, policymakers, and practitioners may benefit greatly from the paper's thorough evaluation of the current literature and identification of potential improvement areas in order to improve data security and safeguard people's privacy.

The literature on AI and ML applications for data protection, including fraud detection, identifying security breaches, and encrypting sensitive data, is reviewed in the second part.

Literature Review

The development and use of AI and ML methods to improve data security and privacy is the focus of the field of artificial intelligence and machine learning in data protection.

To discover security breaches, detect fraud, and encrypt sensitive data, one method is to employ AI and ML algorithms. This method has produced encouraging outcomes in a number of applications, including intrusion detection and credit card fraud detection. The disadvantage of this method is that it needs a lot of data to train the algorithms, and sometimes getting that data might be difficult.

Another strategy is to employ privacy-preserving methods like homomorphic encryption and differential privacy. With the use of these methods, sensitive data may be analysed while maintaining individuals' privacy. This method has the disadvantage that it can be computationally expensive to implement and may need a lot of resources.

A more integrated strategy that includes the advantages of all these techniques is emerging as the winner of the conflict between them. For instance, privacy-preserving strategies may be used to improve the performance of AI and ML algorithms, and AI and ML algorithms can be trained using sensitive data while yet maintaining the privacy of that data.

The three main unresolved issues in the discipline are algorithmic bias, ethical issues, and privacy issues. Addressing these issues and creating more moral and open algorithms are necessary for the responsible and transparent use of AI and ML in data protection. Further study is required to examine novel data protection applications for these technologies and to source prospective areas for development.

Artificial intelligence and machine learning have been used in the past to protect data in a variety of ways that improve security and privacy. Machine learning algorithms have been employed, for instance, in the identification of credit card fraud to examine past transaction data and spot patterns of fraud. Based on their content and other factors, incoming emails are classified as spam or not using AI algorithms for spam filtering.Additionally, machine learning techniques have been developed for intrusion detection systems to detect unusual activity that may point to a security breach. These algorithms have been trained using network traffic data. These strategies do have some drawbacks, though, including the requirement for a lot of data to train the algorithms, the danger of algorithmic bias, and the possibility for privacy violations. In order to overcome these difficulties, the study suggests an integrated strategy that leverages the advantages of all of these approaches. The purpose of the study is to emphasize the need for a more responsible and ethical use of AI and ML in data protection and to offer an integrated strategy that solves the shortcomings of current approaches.

The approaches and systems employed in the field of artificial intelligence and machine learning for data protection have an underlying theory that is based on mathematical models and notions.

1. Modelling data uncertainty and calculating the likelihood of various outcomes are both done using probability theory. In machine learning, the likelihood of various

outcomes is modelled using the theory of probability given the input data. For instance, a machine learning system may be trained on a dataset of historical transaction data in order to detect credit card fraud. Each transaction in this dataset is either flagged as fraudulent or not. With the use of this information, the computer can create a probabilistic model that forecasts the likelihood of fraud in each new transaction.

2. Data analysis and forecasting based on statistical models both include the application of statistical methods. To estimate a model's parameters and assess a model's performance on fresh data, machine learning uses statistical approaches. A machine learning system, for instance, may be trained on a dataset of emails that have been classified as spam or not spam, with each email having been given a label. Based on the email's content and other factors, the program utilizes statistical techniques to develop a model that predicts the likelihood that it is spam.
3. By minimizing or maximizing an objective function, optimization theory is used to determine the optimum solution to a given issue. In machine learning, optimization theory is used to identify the model's parameters that minimize the discrepancy between expected and actual results. For instance, in intrusion detection systems, a machine learning algorithm may be trained on a dataset of network traffic data, where each data point is tagged as normal or abnormal. The approach applies optimization theory to learn a model that minimizes the discrepancy between the expected and actual labels for fresh network traffic data.

In general, the underlying philosophy of these methods and systems is founded on mathematical ideas and models that allow for the analysis and manipulation of data to enhance data.

Methodology

A review and study of the literature served as the research methodology for this work. The authors did an extensive assessment of prior studies in the field of artificial intelligence and machine learning in data protection, evaluating the benefits and drawbacks of various strategies. Additionally, the authors gave examples to highlight the key ideas and methodologies employed in this area, showing how probability theory, statistical approaches, and optimization theory are utilized in various contexts of AI and ML in data protection. A critical assessment of the ethical and societal ramifications of using AI and ML for data protection is also included in the report, including concerns around algorithmic bias, privacy, transparency, and responsibility.

Results

This paper offers a thorough and critical overview of the state of research in the area of artificial intelligence and machine learning in data protection, highlighting the advantages and disadvantages of current strategies and suggesting an integrated and responsible strategy to address the issues and implications of this rapidly developing field.

Murakonda and Shokri (2020) indicated that sophisticated tools such as ML privacy meters can be used for evaluating the privacy risk of the data which are being used for machine learning models. ML-based tools are also used for analyzing compliance issues with data protection regulations. The pregnancy meter is capable of quantifying the privacy risk to the training data. It can also use algorithms for evaluating the privacy risk of machine learning models through the inference attack. The privacy risk score offered by the tool is useful for identifying vulnerable data records. The model is capable of estimating the amount of information that can be accessed by hackers through prediction. The tool is also useful for identifying the potential threats to the training data. It produces detailed privacy reports by comparing and contrasting the different data classes.

Aung et al., (2021) also indicated the applicability of AI and ML-based solutions to ensure network security in the Healthcare industry. The devices used in this industry deal with highly sensitive data regarding the personal information and health status of individuals. The use of AI and ML-based techniques in data pre-processing, feature engineering, model optimization, and evaluation would be effective to comply with the national rules and regulations and enhance data security.

Haider et al., (2020) indicated the threats of network security are high due to the latest advancement of the 5G networks. The 5G networks can be characterized by extensive softwareization, virtualization, and cloudification. The Rapid evolution of Network Technology also increased the risk of unauthorized access. Artificial Intelligence and machine learning place an important role in modelling, designing, and implementation of the security protocols. The protocols are effective to protect the network from a wide range of threats. Artificial intelligence animation learning can support the development of data-driven decisions while managing the virtualized network elements. The data is transferred through multiple layers in the highly software-centric 5G networks. Depending on the nature of the network, the security levels can vary. The rapid advancements in artificial intelligence and machine learning can be useful for the robustness of the 5G networks by improving privacy, security, and threat-detection capabilities.

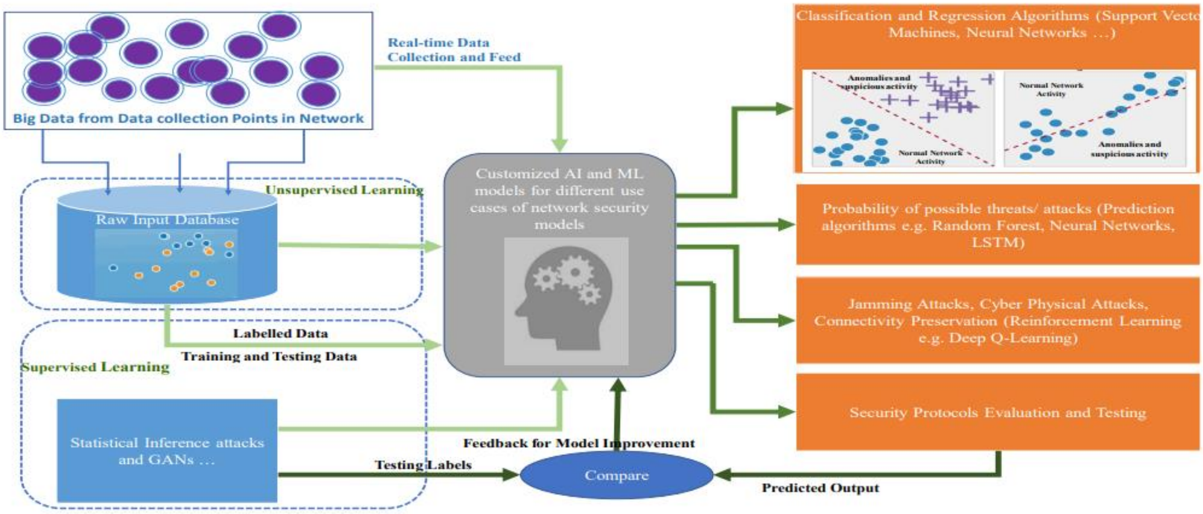
Big Data from Data collection Points in Network
Real-time Data Collection and Feed
Unsupervised Learning
Raw Input Database
Labelled Data
Training and Testing Data
Supervised Learning
Statistical Inference attacks and GANs ...
Customized AI and ML models for different use cases of network security models
Feedback for Model Improvement
Testing Labels
Compare
Predicted Output
Classification and Regression Algorithms (Support Vecto Machines, Neural Networks ...)
Anomalies and suspicious activity
Normal Network Activity
Normal Network Activity
Anomalies and suspicious activity
Probability of possible threats/ attacks (Prediction algorithms e.g. Random Forest, Neural Networks, LSTM)
Jamming Attacks, Cyber Physical Attacks, Connectivity Preservation (Reinforcement Learning e.g. Deep Q-Learning)
Security Protocols Evaluation and Testing

Figure 1- Different Applications of AI and ML

(Source- Haider et al., 2020)

The classification algorithms are used for detecting anomalies by reviewing the network parameters with the help of throughput and network logs. The clustering algorithms are useful for categorizing different types of threats and loopholes in network security. The models such as the generative adversarial network and statistical importance attacks can be used for examining the performance of new security measures. AI and ML models can improve encryption, secure computation and promote hybrid learning. However, the performance of the AI and ML tools can be optimized by reviewing the authentication, access control, authorization, and privacy-preserving strategies. Eavesdropping is another key concern in network security. AI and ML applications also need to focus on mitigating the eavesdropping risk.

Discussions

Cybersecurity might be revolutionized by the use of AI and ML to data protection techniques, an area that is quickly developing. There are several advantages to employing AI and ML in data protection since these technologies can analyze vast volumes of data rapidly and correctly, making it simpler to identify and address security problems in real time.

The possibility of bias in AI and ML algorithms is another difficulty. An algorithm may be biased if it was developed on biased data or if it has biases of its own. Data protection techniques may be complicated if this leads to unjust or discriminatory effects.

Making ensuring AI and ML algorithms are developed and applied in a transparent, explicable, and responsible manner is crucial in order to overcome these difficulties. This might involve adopting ethical standards for AI and ML, employing open-source methods, and offering explicit documentation.

Overall, the incorporation of AI and ML into data protection techniques has the potential to greatly improve cybersecurity measures and shield sensitive data from ever-changing threats. However, security measures need to be strengthened by incorporating new AI and ML-based techniques. The emphasis should be given to addressing new threats such as eavesdropping and authorization issues.

Conclusion

The volume of data collected and exchanged on digital platforms, as well as the sophistication of cybersecurity threats, are both significant concerns in today's digital age, making data protection a crucial issue. Sensitive data must now be protected using modern data protection techniques that can adapt to changing threats and guarantee the confidentiality, integrity, and availability of data. Traditional security measures are no longer adequate to do this. Overall, applying AI and ML to data protection techniques such as secure computation, and analysis of network traffic have the potential to significantly improve cybersecurity measures and shield sensitive data from changing threats, but it's important to be aware of the difficulties and constraints posed by these technologies and adopt a comprehensive approach to data protection.

References

Artificial Intelligence (AI) for Cybersecurity | IBM. (n.d.). https://www.ibm.com/security/artificial-intelligence?utm_content=SRCWW&p1=Search&p4=43700074604519848&p5=p&gclid=CjwKCAjw9J2iBhBPEiwAErwpeVZ_PnlF59tsuPuf7ICqZQuf1q6nDFx8I0fRtVbmGUhc-7f2kiGQCxoCKpwQAvD_BwE&gclsrc=aw.ds

Murakonda, S. K., & Shokri, R. (2020). ML Privacy Meter: Aiding regulatory compliance by quantifying the privacy risks of machine learning. *arXiv preprint arXiv:2007.09339*.

Haider, N., Baig, M. Z., & Imran, M. (2020). Artificial Intelligence and Machine Learning in 5G Network Security: Opportunities, advantages, and future research trends. *arXiv preprint arXiv:2007.04490*.

Aung, Y. Y., Wong, D. C., & Ting, D. S. (2021). The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. *British medical bulletin*, *139*(1), 4-15.

Kerry, C. F. (2022, March 9). Protecting privacy in an AI-driven world. *Brookings*. https://www.brookings.edu/research/protecting-privacy-in-an-ai-driven-world/