International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

xx

COPY RIGHT

IJIEMR Transactions, online available on 19th Jun 2024. Link

https://www.ijiemr.org/downloads/Volume-13/ISSUE-6

10.48047/IJIEMR/V13/ISSUE 06/04

TITLE: Leveraging AI for PII Identification and Data Masking: Techniques and Benefits

Volume 13, ISSUE 06, Pages: 26-31

Paper Authors **: SAI CHARAN DATTA**

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Leveraging AI for PII Identification and Data Masking: Techniques and Benefits

SAI CHARAN DATTA
EMAIL – CHARAN.DATTA@GMAIL.COM
COMPANY – SYSTEM SOFT TECHNOLOGIES

Abstract—This study highlighted the use of artificial intelligence (AI) for detecting PII along with data masking with a key reference to the techniques and benefits associated with the procedures. It has been identified that PII and data masking related processes are often shared or rather sold to other companies which often creates a scope for the hackers to use these data for identity theft, selling of data, or ransom ware attacks. Enhance Encryption Standards support in enhancing integrity, security of sensitive information. However, threats of cyber security are a major challenge with PII identification for protecting crucial information. Therefore, it is recommended to integrating policies while using AI-integrated PII identification and data masking.

Keywords—PII, AI, Cybersecurity

I. Introduction

Background

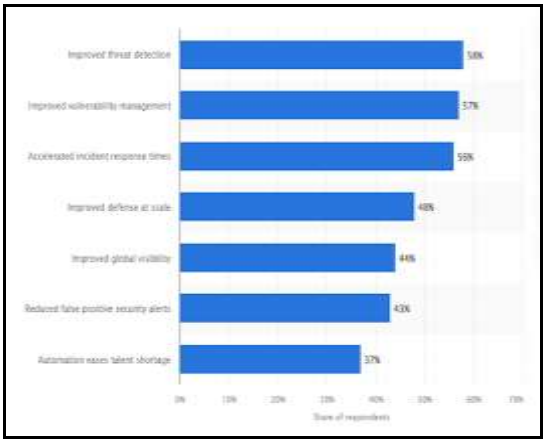

Figure 1: Main benefits of incorporating AI into cybersecurity operations 2023

Personally Identifiable Information (PII) detection is among the key features offered by artificial intelligence (AI) within the cloud which helps in developing key and advanced applications which include written language. As per figure 1, a survey of 2023 reported that around 60% of members worldwide agreed on the capacity to use artificial intelligence to improve threat detection in their cybersecurity activities [1]. The feature of PII detection is being able to recognise, categories, and redact sensitive data within an unstructured text. The dual-ledged pattern of AI in terms of PII has been offering unprecedented data insights as the AI-driven analytics can positively recognise trends, personalise consumers' services and estimate market movements [2]. In context of AI, producing and processing huge amounts of information, including PII, have become highly relevant that enables it to recognise an individual through their names, addresses to phone numbers, passport data and other security numbers. On other hand, data masking refers to the key process of hiding information through modifying its real letters and the numbers. However, as of 2023, around 52% of all information breach in global companies included consumer personal identifiable (PII) data which led their businesses to face loss of credentials, breach and other security issues [3]. Hence, AI has been evidently significant in PII identification and data masking which also helps initiating this present study.

Problem statement

The identified problem in this study is that PII and data masking related processes are often shared or rather sold to other companies which often creates a scope for the hackers to use these data for identity theft, selling of data, or ransomware attacks.

Aim

The aim of this study is to highlight the use of artificial intelligence (AI) for detecting PII along with data masking with a key reference to the techniques and benefits associated with the procedures.

Objectives

●To identify the key techniques of using AI in PII identification and data masking

●To recognise key factors related to AI that influence the identification of PII and data masking and benefits associated with it

●To highlight major challenges faced by companies while leveraging AI for PII identification and data masking

●To recommend strategies for overcoming the identified problems related to the utilisation of AI in identification of PII and data masking

Questions

●What are the key techniques of leveraging AI in PII identification and data masking?

●What are some major factors related to AI, influencing the identification of PII and data masking and benefits associated with the technology?

●What are the common challenges encountered by companies while leveraging AI for PII identification and data masking?

●What strategic measures can be applied for reducing the challenges related to the use of AI in identification of PII and data masking?

II. Literature Review

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Techniques of using AI in PII identification and data masking

AI has been given a large amount of appreciation and recognition from the companies in varied sectors for making their businesses able to reduce security issues through key identification of PII in terms of theft recognition and resolving data masking issues as well. Incorporating powerful tools like AI in terms of managing and analysing PII, the security system of any company can be significantly improved like an automated part of data collection and utilisation processes. One of the key practices or techniques related to the leveraging of AI in PII identification is implementing an encryption solution. Encryption offers another stage or layer of protection in cloud environments, through encoding the information both in transit, while at ease [4]. Companies need to use AI for either PII or data masking in a way that helps them in reducing their security issues, rather than making them more vulnerable or exposed to hackers. However, data hiding techniques or methods protect the sensitive information through concealing it as it is in a multimedia file in a way which does not increase any suspicions of any sensitive information [5]. Hence, both the techniques to manage PII detection and data masking in relation to AI have been assessed to be significant in ensuring higher security for companies.

Factors and benefits associated with the application of AI in identification of PII and data masking

Key factors related to the use of AI in managing PII detection and data masking include ethical practices in the use of AI, performance expectancy, technical knowledge and others. For example, like all other ethics related to the digital sector, AI ethics is also highly interdisciplinary as well as is rapidly developing [6]. AI ethics help companies in terms of ensuring that the technology is developed as well as utilised in an accountable way which indicates a secure, safe, humane, and environmentally friendly attitude towards AI. On other hand, key benefits related to the utilisation of AI in terms of PII identification and data masking include data protection, personalised experiences, advanced language generation models, and data security control mechanisms. In context of data masking, real data is changed as there is no reversibility whereas in terms of data encryption, the real data is shifted to encoded data and restored by applying a key. Thus, the factors and benefits identified above highlight that AI has relevant advantages in detecting PII and data masking.

Challenges faced by companies to use AI for PII identification and data masking

Challenges related to the application of AI in identifying PII and managing data masking include poor data confidentiality maintenance, data vulnerability, exposure to hackers, malware attacks, data selling, identity theft and others. For instance, identity theft has been one of the initial stages in a two-stage process through which particular kinds of personally identifiable information (PII) are stolen as well as applied for committing identity theft [7]. People often have the information which need to be secured as well as need to take care of guarding their PII accordingly. Hence, the challenges of data breach, poor confidentiality, and identity theft led companies to face challenges in terms of using AI in preventing PII and data masking.

Strategies to resolve identified problems related to the use of AI in identification of PII and data masking

Key strategies related to leveraging AI in terms of PII detection and data masking include encryption, redaction, data anonymization, formulation of data safekeeping data policies, identification of precise PII, identification of sensitive data, scrambling and others. For instance, data anonymization has been another core principle which engages in the output assessment as the used technology is related to obtaining the required process [8]. Additionally, "transparency, consent, data minimization, and anonymization" have been recognised as key contexts for gaining privacy. Hence, these are some of the strategies identified to resolve identified problems related to the use of AI in identification of PII and data masking.

Theoretical Perspective

Technology acceptance model

Figure 2: Technology acceptance model

"Technology acceptance model (TAM)" defines that the perceived ease of use and perceived usefulness of technology lead to informed decision making related to the acceptance of one or more technologies [9]. Application of TAM might significantly help the companies that are seeking to offer an AI-driven cyber security solution for identifying PII and data masking. TAM is a key information process theory which models the ways in which users come to accept as well as apply a technology [10]. Hence, implementation of TAM would significantly help companies in improving their approach of using AI in PII detection and data masking.

Conceptual framework

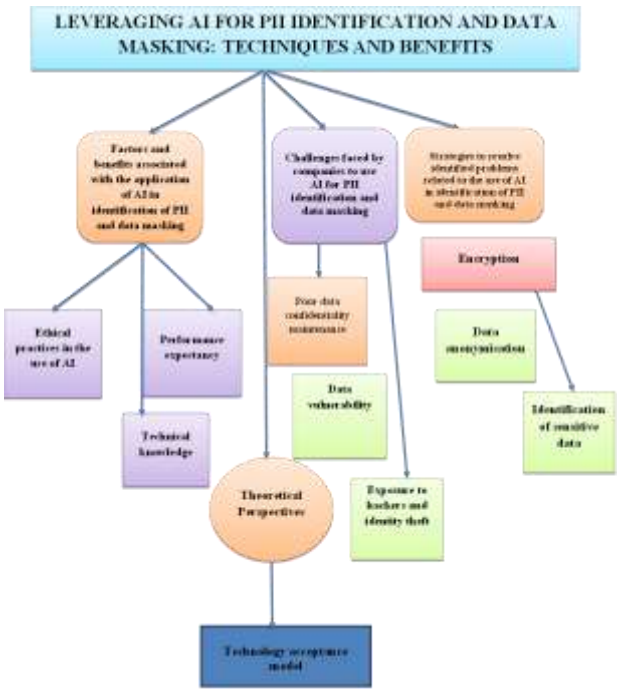

Figure 3: Conceptual framework

(Source: Created by author)

Literature gap

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Assessing existing literature on the use of AI in PII detection and data masking, this study has recognised some of the key gaps in those literature works. Some of the commonly identified gaps in the studies included lack of an integrated approach towards PII and data masking, poor focus on AI as a key technology to resolve cybersecurity issues. Hence, this study has focused on fulfilling these gaps through interpreting relevant data in a systematic and clearer way.

III. Methods

Research Onion

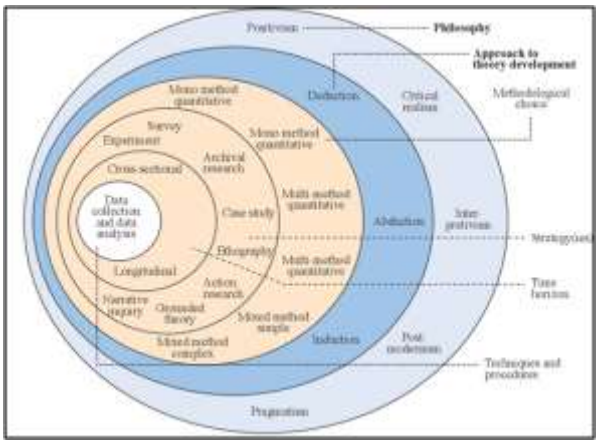

Figure 1: Research Onion

Research onion was followed for constructing the road map for methodology in which "research philosophy, research approach, data collection method, and data analysis" demonstrated in the following section.

Research philosophy

"Interpretivism research philosophy" was adopted within the research in order to gain a subjective view of the research. Interpretivism philosophy is more concerned with in depth evaluation of human factors along with physical phenomenon [14]. Considering this, PII identification or data masking by leveraging AI reflects human knowledge, perception regarding using advanced technology. By applying, interpretivism standpoint this could effectively evaluate factors like human skill as it considered human factor and capability of AI in PII identification. On the other hand, "positivism philosophy" would not be able to provide an illustration of the human factor that might create a knowledge gap during the research. Thus, "interpretivism philosophy" was adopted to gain a holistic view by considering human factors to conduct research on leveraging AI for PII identification and data masking.

Research approach

"Inductive research approach" was adopted for gaining a comprehensive picture on the cause-and-effect relation regarding PII identification. In addition, an "inductive approach" uses evidence for evaluating a phenomenon which supports in reaching a sound conclusion. "Inductive approach takes a specific case and draws a general concept for the case [15]. On the other hand, the "deductive approach" puts less emphasis upon evidence for evaluating the phenomenon PII identification that might lead to wrong conclusions. Therefore, "inductive approach" was utilising for increased authenticity of the research as "inductive approach" uses evidence.

Research strategies

"Archival research strategy" was used for collection of relevant information that addresses research questions. "Archival research strategy" supports the collection of a wide range of data sets associated with PII information, data masking, and identification of PII through AI.

Data collection method

"Secondary qualitative data collection" method was applied for conducting the research on leveraging AI for data masking and PII identification. Secondary data is cheaper and easier way to obtain comparison with primary data [16]. On the other hand, primary data collection needs a huge amount of time to interact with participants. Thus, "secondary data collection" was used for accumulating relevant information from a wide range of resources in a cost-effective manner.

Search strategy

Keyword Based search strategy was applied for collecting relevant journals on the PII identification and data masking by leveraging AI. "PII, data making, data privacy law, AI technology, and social security" was used as keywords for filtering out crucial journals relevant to research questions.

Inclusion and exclusion criteria

This study included the use of journal articles that have been published since 2020, written in English language with a precise structure and content in the documents. Articles related to the use of AI, PII detection, data hiding or masking, and theories on technological execution have been considered in this study. Lastly, journal articles that only include an abstract without a full concise analysis, published before 2020, written in other languages except English were discarded from this research.

Data analysis

"Thematic analysis" was adopted for finding patterns between research variables in order to find the relation between AI adaptation in PII identification and data masking. Thematic analysis plays a substantial role in identifying patterns across dataset [17]. Referencing this, the relationship between data pattern like benefits of data masking, leveraging AI assist in PII identification. Fours themes were development for evaluating the research topics from different perspectives like influencing factors, challenges, strategies associated with using AI for PII identification and data masking. Therefore, "thematic analysis" was used for performing the research as it identifies patterns and eases the process of accomplishing research objectives.

Ethical consideration

Ethics in research becomes an important element to maintain transparency, confidentiality, integrity and other key ethical guidelines [11]. Some of the key ethics maintained in this secondary qualitative research included "respect for autonomy, data protection regulation, beneficence, non-maleficence, informed consent, confidentiality, and justice". Like a universal principle, the concept of autonomy needs to be included in research for receiving the warrant a right for carrying out the study [12]. Hence, adhering to these above-mentioned ethical guidelines significantly helped this research in meeting its goals accordingly and ethically.

IV. Data analysis

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Theme 1: The techniques of Encryption standards used for leveraging AI in PII identification and data masking

The incorporation of artificial intelligence (AI) reflects the revolutionary changes for the institution in order to manage "personalised identified information" (PII) due to the large capacity of AI. Development of the more secure data algorithm in AI is ongoing in which "Enhance Encryption Standards" aims for improving integrity, security of financial data [18]. Referencing this, integration of AI-enabled "enhanced encryption standard" techniques safeguarding data associated with PII security. PII involves information sensitive information like "passport number, social security number, credit card number" which needs greater protection. Therefore, AI enabled systems utilises techniques like "enhanced encryption standard" for the identification of PII and data masking in a secure way.

Theme 2: Protection of data is the major benefit of leveraging AI for PII identification and data masking

Advancement of technologies like cloud storage eases the process of data collection and alleviates the needs for paperwork's. PII holds sensitive information of customers in which a poor emphasis on the factors could result in major financial loss for customers. An organisation maintains the regulation of GDPR while managing information related to PII to protect the privacy of sensitive information. In this regard, a chatbot application, "Replika" allows the opportunity to share users' feelings more than just PII by the GDPR definition [19]. Considering this, utilisation of GDPR rule could be integrated with AI for maintenance of privacy related to PII identification. Additionally, AI integration eases the process data masking by replacing sensitive information with false data. Therefore, integration of AI in PII identification and data masking provide benefits regarding protection of sensitive data.

Theme 3: Breach of data is a major challenge associated with AI in PII identification and data masking

Data breach is considered to be a major challenge associated with PII identification and data masking. AI-enabled systems automate the tasks that raise concern for security issues due to poor maintenance of the system. The type of PII shared during transaction, correlation, retention time and sensitivity also severely affect privacy risk levels [20]. Data masking could potentially help in protection of PII data; however it raises concern regarding data breaches through malware attack. For instance, 146 million people were affected due to data breaching at "Equifax " and how alarming the problem is regarding PII data breaches [21]. Therefore, data breaching incidents of PII information become a major challenge for organisations.

Theme 4: Data safekeeping policies could be used as effective strategy for protecting sensitive information associated with AI enabled PII identification and data masking

Integration of effective policies supports the reduction of unnecessary access to the PII information that eliminates the threats of data breach incidents. "Access control policy" applied by system administrators supports restricting user ability regarding modification of accessed permission [22]. On the other hand, Data masking methods assists in replacing sensitive PII information with anonymous data for protecting privacy. Therefore, malware attack or unauthorised interaction for PII identification through leveraging AI could be avoided by integrating data safekeeping policies.

V. Discussion

PII holds several sensitive information such as driver licence, financial information and medical records. Considering this, integration of AI eases the process of finding necessary information related to an individual from a vast number of sources. "Encryption solution" technique is used during finding necessary data through leveraging AI. In addition, integration AI in finding necessary data also ensures safeguarding sensitive information of people regarding PII. It emphasises the compliance with GDPR rules for the protection of personal information while incorporating AI for PII finds and data masking. However, identity theft is considered to be a major challenge associated with PII [7]. In this regard finding section highlights that PII shred during transaction, correlation could severely increase the risk level. Extraction of PII by itself like a personal address could pose personal threats [24]. Hence, "accessed control policy" could be implemented as an effective strategy in order to restrict users' control [22]. Additionally, the data masking process also supports releasing sensitive information with anonymous information to protect PII data and gaining Differential privacy. Differential privacy is considered to be popular privacy today due to strengthening the security to any changes to data set [25]. Therefore, restricting user access, safekeeping policy compliance assists in project PII information while leveraging AI for PII identification and data making.

VI. Contribution and Limitation

AI assists in automating the task in identifying PII data of an individual. This report contributes in demonstrating knowledge on the benefits of AI during identification of PII and data masking. Additionally, the research also assists in demonstrating challenges associated with AI integration during PII identification. This supports organisations, individuals and governments to adopt policies to close the loopholes. Apart from this, the study is focused on overall information regarding PII identification through leveraging AI. Therefore, a lack of emphasising any single PII data is considered to be a major limitation for the research.

VII. Conclusion and recommendation

From the above discussion it can be concluded that employee AI integration eases the process of PII identification and data masking from a vast number of sources. Utilisation of encrypted solutions supports identification of data by leveraging AI in a secure manner. Additionally, AI integration along with GDPR policies assures protection of PII information of individuals. However, it still raises challenges for data breaches due to improper maintenance of AI for identifying PII. Referencing this, integration of policies assists to restrict the user's access to PII information during search process that alleviates chances regarding data breaches.

Recommendation

Utilisation of homomorphic encryption technology for protecting AI-integrated PII identification

Alleviation of the threats regarding unencrypted data potentially help in improving data security during AI enabled PII identification to address data breach issue. The neural network model of "homomorphic encryption "assists in prediction of the encrypted data [23]. In this regard, identification of unencrypted data during AI-integrated PII identification streamlines the process of data protection. Therefore, it is recommended to utilise

homomorphic encryption technology supported in identification of such encrypted data to shield crucial information of individuals.

REFERENCES

[1] S. O. Olabanji, O. B. Oladoyinbo, C. U. Asonze, T. O. Oladoyinbo, S. A. Ajayi, and O. O. Olaniyi, "Effect of Adopting AI to Explore Big Data on Personally Identifiable Information (PII) for Financial and Economic Data Transformation," SSRN Electronic Journal, 2024, doi: 10.2139/ssrn.4739227.

[2] Petrosyan, A., "Topic: Data breaches worldwide," Statista, Oct. 25, 2023. https://www.statista.com/topics/11610/data-breaches-worldwide/#topicOverview (accessed Jun. 07, 2024).

[3] A. Borgeaud, "Top benefits of integrating AI into cybersecurity 2023," Statista. https://www.statista.com/statistics/1425575/top-benefits-of-incorporating-ai-into-cybersecurity-operations/ (accessed Jun. 07, 2024).

[4] P. Desai and T. Hamid, "Best Practices for Securing Financial Data and PII in Public Cloud," International Journal of Computer Applications, vol. 183, no. 40, pp. 1–6, Dec. 2021, doi: 10.5120/ijca2021921737.

[5] F. S. Hassan and A. Gutub, "Efficient reversible data hiding multimedia technique based on smart image interpolation," Multimedia Tools and Applications, vol. 79, no. 39–40, pp. 30087–30109, Aug. 2020, doi: 10.1007/s11042-020-09513-1.

[6] E. Kazim and A. S. Koshiyama, "A high-level overview of AI ethics," Patterns, vol. 2, no. 9, p. 100314, Sep. 2021, doi: 10.1016/j.patter.2021.100314.

[7] N. L. Piquero, A. R. Piquero, S. Gies, B. Green, A. Bobnis, and E. Velasquez, "Preventing Identity Theft: Perspectives on Technological Solutions from Industry Insiders," in The New Technology of Financial Crime, London: Routledge, 2022, pp. 163–182. Accessed: Jun. 07, 2024. [Online]. Available: http://dx.doi.org/10.4324/9781003258100-9

[8] B. Dash, P. Sharma, and A. Ali, "Federated Learning for Privacy-Preserving: A Review of PII Data Analysis in Fintech," International Journal of Software Engineering & Applications, vol. 13, no. 4, pp. 1–13, Jul. 2022, doi: 10.5121/ijsea.2022.13401.

[9] E. S. Vorm and D. J. Y. Combs, "Integrating Transparency, Trust, and Acceptance: The Intelligent Systems Technology Acceptance Model (ISTAM)," International Journal of Human–Computer Interaction, vol. 38, no. 18–20, pp. 1828–1845, May 2022, doi: 10.1080/10447318.2022.2070107.

[10] H. Alqahtani and M. Kavakli-Thorne, "Factors Affecting Acceptance of a Mobile Augmented Reality Application for Cybersecurity Awareness," in Proceedings of the 2020 4th International Conference on Virtual and Augmented Reality Simulations, Feb. 2020. Accessed: Jun. 07, 2024. [Online]. Available: http://dx.doi.org/10.1145/3385378.3385382

[11] J. Bos, Research Ethics for Students in the Social Sciences. Cham: Springer International Publishing, 2020. Accessed: Jun. 07, 2024. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-48415-6

[12] A. Traianou and M. Hammersley, "Is there a right not to be researched? Is there a right to do research? Some questions about informed consent and the principle of autonomy," International Journal of Social Research Methodology, vol. 24, no. 4, pp. 443–452, Aug. 2020, doi: 10.1080/13645579.2020.1801276.

[13] H. Arbale and D. N. Mutisya, "Book Review: 'Research Methods for Business Students' (Eighth Edition) by Mark N. K. Saunders, Philip Lewis, and Adrian Thornhill (Pearson Education, 2019)," African Quarterly Social Science Review, vol. 1, no. 2, pp. 8–21, Apr. 2024, doi: 10.51867/aqssr.1.2.2.

[14] M. Junjie and M. Yingxin, "The Discussions of Positivism and Interpretivism," Global Academic Journal of Humanities and Social Sciences, vol. 4, no. 1, pp. 10–14, Jan. 2022, doi: 10.36348/gajhss.2022.v04i01.002.

[15] H. Taherdoost, "Data collection methods and tools for research; a step-by-step guide to choose data collection technique for academic and business research projects," International Journal of Academic Research in Management (IJARM), 10(1), pp.10-38. 2021.

[16] C. Okoli, "Inductive, abductive and deductive theorising," International Journal of Management Concepts and Philosophy, vol. 16, no. 3, pp. 302–316, 2023, doi: 10.1504/ijmcp.2023.131769.

[17] K. L. Peel, "A Beginner's Guide to Applied Educational Research using Thematic Analysis," Practical Assessment, Research, and Evaluation, vol. 25, no. 1, 2019, doi: 10.7275/ryr5-k983.

[18] S. O. Olabanji, O. B. Oladoyinbo, C. U. Asonze, T. O. Oladoyinbo, S. A. Ajayi, and O. O. Olaniyi, "Effect of Adopting AI to Explore Big Data on Personally Identifiable Information (PII) for Financial and Economic Data Transformation," SSRN Electronic Journal, 2024, doi: 10.2139/ssrn.4739227.

[19] M. Hasal, J. Nowaková, K. Ahmed Saghair, H. Abdulla, V. Snášel, and L. Ogiela, "Chatbots: Security, privacy, data protection, and social aspects," Concurrency and Computation: Practice and Experience, vol. 33, no. 19, Jun. 2021, doi: 10.1002/cpe.6426.

[20] P. Silva, C. Gonçalves, N. Antunes, M. Curado, and B. Walek, "Privacy risk assessment and privacy-preserving data monitoring," Expert Systems with Applications, vol. 200, p. 116867, Aug. 2022, doi: 10.1016/j.eswa.2022.116867.

[21] A. K. Makhija, "Deep Learning Application – Identifying PII (Personally Identifiable Information) to

Protect," Journal of Accounting, Finance, Economics, and Social Sciences, vol. 5, no. 2, pp. 10–16, 2020, doi: 10.62458/jafess.160224.5(2)10-16.

[22] Oluwatoyin Ajoke Fayayola, Oluwabukunmi Latifat Olorunfemi, and Philip Olaseni Shoetan, "DATA PRIVACY AND SECURITY IN IT: A REVIEW OF TECHNIQUES AND CHALLENGES," Computer Science & IT Research Journal, vol. 5, no. 3, pp. 606–615, Mar. 2024, doi: 10.51594/csitrj.v5i3.909.

[23] P. Mohassel and Y. Zhang, "SecureML: A System for Scalable Privacy-Preserving Machine Learning," in 2017 IEEE Symposium on Security and Privacy (SP), May 2017. Accessed: Jun. 07, 2024. [Online]. Available: http://dx.doi.org/10.1109/sp.2017.12

[24] N. Lukas, A. Salem, R. Sim, S. Tople, L. Wutschitz, and S. Zanella-Béguelin, "Analyzing Leakage of Personally Identifiable Information in Language Models," in 2023 IEEE Symposium on Security and Privacy (SP), May 2023. Accessed: Jun. 07, 2024. [Online]. Available: http://dx.doi.org/10.1109/sp46215.2023.10179300

[25] Ji, S., Lipton, Z. C., Li, X., & A. Anandkumar, "Differential Privacy and Machine Learning: a Survey and Review". arXiv preprint arXiv:1907.04064.2019.