Impact of machine learning techniques in Privacy compliance

Pramod Misra[1]*, Neha Tiwari[2]

[1] Department of Computer Science, Georgia Institute of Technology, Atlanta, USA

[2]Department of Computer Science, Bansal Institute of Science and Technology, Bhopal, India

Correspondence email- pmisra30@gatech.edu

Abstract

In today's era, "artificial intelligence" (AI) and "machine learning" (ML) are integral to our daily life and to find solutions to complicated problems under a variety of conditions, capabilities we previously believed to be exclusive to humans. To increase its usage, most of our tasks are depended on AI and ML and are solved by them. Personal information, financial information, health information, etc are also shared, stored and processed by AI. With the advent of this, privacy became a major issue, which is believed to be solved by AI aided by ML. Furthermore, the protection of data privacy and machine learning (ML) in the beginning stage as well as the previous resolutions are the main priority on the privacy issues during the machine learning methods. This study is intended to provide a deep understanding of the importance of privacy protection problems and the application of machine learning in this regard.

Introduction

The terms "artificial intelligence" (AI) (Greener et al., 2022) and "machine learning" (ML) is used to refer to computer systems that can learn from their own mistakes and find solutions to complicated problems under a variety of conditions, capabilities we previously believed to be exclusive to humans. And data (European Commission, 2020), oftentimes personal data, is what powers these systems, allowing them to grow clever and learn. Recent years have seen significant progress in AI research, and the future looks bright: a stronger and more effective public sector, new approaches to environmental and climate change mitigation, a safer society, and perhaps even a cancer cure. Our privacy rights will be strengthened by the new data protection laws, which take effect in May 2018 and impose stricter restrictions on businesses that process such data (Tabassi et al., 2019). As a result of the policy, organizations will be held more accountable for processing personal data in compliance with the rules, and transparency standards will be stricter. Demand for data is increasing at the same time as requirements are becoming more stringent. Only when there is an abundance of pertinent data for them to learn from will AI-based systems become intelligent (Janiesch et al., 2021).

Data privacy risks of machine learning models

Massive volumes of personal data are being gathered by businesses to create machine learning-based apps. This information is utilized to train the models frequently including private data on specific people (Janiesch et al., 2021). Models of machine learning encode data about the training datasets that they are using. The intended purpose of the encoded data is to reflect the broad trends that underlie the demographic statistics. However, it is frequently noticed. That these models can memorize details about some either be fooled into doing so or be members of their training data (European Commission, 2020).

Deep neural networks and other models with high capacity and a high generalization gap are more likely to memorize data items from their training set. The model's predictions, which behave differently on training data compared to test data, and its parameters, which hold statistically correlated information about particular data points in their training set, both reflect this (McMahon B et al., 2017). Using membership inference attacks, which allow an attacker to identify the presence of a certain record in a model's training dataset just by watching the model, this flaw in machine learning models was demonstrated. Both in the black-box and white-box environments, it was demonstrated that machine learning models are vulnerable to these attacks (Goldsteen et al., 2022).

We can only observe the model's predictions in the black-box setting. This environment simulates the scenario of machine learning as a cloud platform service provided by businesses like Amazon, Microsoft, and Google (McMahon B et al., 2017). It can be used to assess the privacy concerns posed by reputable model users who request predictions in response to their questions. We may also view the model's parameters in the white-box environment. This is representative of a situation where a model is shared with an aggregator in a federated learning environment or outsourced to a cloud or server (Goldsteen et al., 2022) that may not be trusted.

The effectiveness of such inference attacks against machine learning models' training data can be used to assess the privacy threats that they pose.

Data Protection Regulations

A quantitative analysis of the privacy concerns posed by these models is essential for ensuring their safe and secure use, as well as ensuring that they do not divulge sensitive data about the training data. Personal data must be protected when utilized in AI systems by data protection laws like the GDPR and AI governance frameworks, and users must have control over their data and knowledge of how it is used (Nasr et al., 2019).

As of Article 35 of the GDPR (Goldsteen et al., 2022), doing a Data Protection Impact Assessment is required for projects employing cutting-edge technology like machine learning (DPIA). Identifying potential data threats and evaluating how they might affect people are the main DPIA procedures. Generally speaking, risk evaluation in DPIA statements emphasizes the danger of security lapses and unauthorized access to the data. By indirectly disclosing information about the training data through the model's predictions and parameters, machine learning methods increase the privacy risk of that data. Data protection laws must therefore receive specific consideration in AI regulatory frameworks (Shokri R and Shmatikov, 2015).

Both the European Commission and the White House have issued guidelines urging the protection of personal data during all stages of the usage of AI systems and the creation of defence-in-depth systems. The Information Commissioner's Office (ICO) for AI audits and the National Institute of Standards and Technology (NIST) for securing applications of Artificial Intelligence both recently released papers that emphasize the risk to data privacy posed by machine learning models (Shokri et al., 2017). Additionally, they particularly point out membership inference as a breach of confidentiality and a potential danger to model training data. Organizations are advised to recognize these vulnerabilities and take precautions to reduce the risk in the auditing framework by ICO. Organizations must account for and estimate the privacy risks to data using models as the ICO's investigative teams will use this framework to judge compliance with data protection rules (Shokri R and Shmatikov, 2015).

ML Privacy meter

Practitioners can benefit from a tool that can automatically assess the privacy issues associated with machine learning models and the training data they use but how do we quantify the chance of unintentional data leaks from sophisticated ML models (Song et al., 2017)? We introduce the ML Privacy Meter, which is based on well-proven methods that measure the privacy risks of machine learning models through membership inference attacks. It can quantify the privacy risks to training data (Murakonda et al., 2020). The tool offers privacy risk scores that assist in identifying the data records that are highly vulnerable to disclosure through model inputs or predictions. At various levels of access to the model, the tool can produce detailed privacy reports regarding the overall and individual risk for data records in the training set. It may calculate the quantity of data that can be obtained by a model's predictions (known as "Black-box access") and its parameters and forecasts combined (known as "White-box access") Consequently, the tool may be used to evaluate the potential risks to training data whether giving query access to the model or revealing the full model (IEEE 21st Conference, 2019).

By employing membership inference attacks against machine learning models, ML Privacy Meter operates (Korba et al., 2007). It replicates a variety of access and model knowledge levels for the attackers. It considers attackers who are limited to using the model's predictions, loss values, and parameters. The tool returns risk scores for all the data records for each of the simulated attacks. These ratings reflect the attacker's perception that the record was included in the training set. The leakage from the model would be greater the difference between the distribution of these scores for records in the training set compared to records outside the training set (Murakonda et al., 2020).

Several studies show that the trade-off between the attacker's False Positive Rate and True Positive Rate can be used to measure the attacker's success. False positive refers to recognizing a non-member as a member while True positive refers to correctly identifying a member as present in the data. An effective attack can produce higher True Positive rates at lower False Positive rates (Song et al., 2017). Equal True Positive and False Positive Rates can be achieved with a simple attack like a random guess. The trade-offs that are accomplished by our simulated attackers are automatically plotted by ML Privacy Meter. The total privacy risk the model poses to the data is quantified by the area under those curves. The risk increases as the area under the curve increases (Aura et al., 2006). These figures can be interpreted as a measurement of information leakage from the model in addition to quantifying the effectiveness of membership inference attacks.

This risk calculation might be helpful when installing machine learning models and doing a Data Protection Impact Assessment. Analyzing, identifying, and minimizing potential dangers to data are the goals of a DPIA (Aura et al., 2006). Practitioners can be guided through all three levels by the ML privacy meter. It can be useful in determining the potential causes of this risk as well as in calculating the privacy risk to data. Additionally, it might help choose and implement effective risk-reduction strategies (Agichtein et al., 2005).

For the training data, the programme generates comprehensive privacy reports. It enables risk comparison between records from various data classes. We may contrast the risk brought on by giving the model black-box access with the risk brought on by white-box access. Practitioners may easily lower the privacy risk by easy steps like fine-tuning their regularization approaches, sub-sampling, re-sampling their data (Aura et al., 2006), etc. because the tool can instantly measure the privacy concerns for training data. Or students can decide to learn with a privacy safeguard in place, such as differential privacy (Kambhatla et al., 2004).

Differential privacy is a concept in cryptography that states that when a single record in the data is changed, the results of the calculation should not be distinguishable. A privacy parameter regulates the degree of indistinguishability. Models with varied privacy assurances can be trained using open-source tools like TensorFlow (Patel et al., 2021) Privacy and Open DP. Choosing a suitable value when using these tools is difficult. Less accurate but greater privacy guarantees are offered by learnt models with a lower value. reflects a privacy risk's worst-case upper bound, whereas the actual risk might be significantly lower. By calculating the risk at each value of epsilon, ML Privacy Meter can assist in the selection of privacy parameters for differential privacy (Han et al., 2003).

Privacy-Preserving Machine Learning (PPML)

Productivity can be greatly increased by using machine learning (ML). The calibre of the data used to train ML models, however, determines how effective ML (Han et al., 2003) systems will be. Additionally, there is a limit to how much data one person or organization can provide when it comes to training ML models. We can unlock value and create potent language models that can be used in a wide range of scenarios, such as text prediction and email reply suggestions, by pooling data to train ML models collectively (Chang et al., 2006). At the same time, we understand the necessity of protecting people's privacy and confidentiality while also gaining and preserving the trust of the users of our products. A key component of our purpose is maintaining the privacy of our client's data. The goal of the Privacy-Preserving in Machine Learning (PPML) programme, which was launched in collaboration with Microsoft and its product teams, is to safeguard customer data privacy and confidentiality throughout the training of large-capacity language models. The PPML programme aims to enhance current methods of securing sensitive information and create new ones that are effective for both individuals and businesses. This contributes to ensuring that the privacy of individuals is protected and that the data is used safely, preventing the leakage of sensitive information (Turmo et al., 2006).

We are currently talking about new research that combines strategies to protect privacy and confidentiality while using critical data to train machine learning models. We show how using PPML may help our ML (Agichtein et al., 2005) pipelines comply with strict privacy rules, and we show that our engineers and researchers have the resources they need to do so. We also go over how PPML best practices help us be open and honest about how we use consumer data (Kambhatla et al., 2004).

A comprehensive strategy for PPML

Recent studies have demonstrated that using ML models might occasionally have unforeseen privacy implications. For instance, very large language models have been demonstrated to memorize training instances, possibly encoding personally identifying information, and pre-trained open language (Sadeghi et al., 2020) models that are fine-tuned on private data might be used inappropriately to recover private information (PII) (Han et al., 2003). Finally, assuming that a particular user was included in the training examples can affect privacy. As a result, we think it's essential to use a variety of strategies to accomplish privacy and confidentiality because no one technique can handle everything by itself. Due to this, we approach PPML using a three-pronged strategy: analyzing the privacy and confidentiality risks and requirements, quantifying the risks, and reducing the likelihood of privacy violations (Duddu and Vasisht, 2018).

Recognize

To help evaluate the secrecy qualities of ML pipelines, we strive to comprehend the threat of customer information leakage and prospective privacy threats. Furthermore, we consider it essential to actively align with decision-makers. We consider national and international regulations governing data privacy, including the General Data Protection Regulation (GDPR) and the European Union's (EU) policy on reliable AI. We then provide tools to explain to policymakers how we adhere to these technological criteria by mapping those legal principles, our contract duties, and responsible AI concepts to our technical requirements (Sadeghi et al., 2020).

Establish

When we are aware of the privacy concerns and the regulations we must follow, we may establish metrics to quantify the risks and monitor our progress in minimizing them (Duddu and Vasisht, 2018).

Mitigate

We create and put into practice mitigation techniques like differential privacy (DP), which is covered in more depth later in this blog article. Following the implementation of mitigation strategies, we evaluate their effectiveness and use the results to improve our PPML strategy (Pitropakis et al., 2019).

Utilizing PPML

We employ the various technologies that go into PPML for a variety of use cases, such as threat modelling and avoiding the leak of training data (Duddu and Vasisht, 2018). For instance, in the following text prediction scenario, we developed quantitative criteria for risk assessment while layering different PPML techniques and taking a comprehensive approach to protecting data privacy.

We just created a tailored assistant for writing emails and documents using the most recent Project Turing natural language (Murakonda et al., 2020) generation models. Based on the present text and other factors, such as the receiver and subject, its transformer-based design uses that to predict the conclusion of a phrase. It is dangerous to use big transformer models since it is possible to memorize and recreate specific training examples while predicting the future, and these examples may contain sensitive data. As a result, we created an approach to both recognise and eliminate sensitive data from the training data. We also made measures to reduce the tendency for memorizing during training. We incorporated PII removal, DP model training, and cautious data selection (Pitropakis et al., 2019).

Preventing the release of private data

Data scientists and Machine Learning (ML) engineers are required to handle client data strictly hands-off as part of our security best practices. However, these mitigations cannot stop more covert privacy breaches, such as the storage of training data in a model that may later be recovered and linked to a user. Because of this, we continue to support the most recent research in this area and use the cutting-edge privacy protections offered by DP (Vorobeychik et al., 2018). Our policies demand a security review, a privacy review, and a compliance evaluation for use cases that influence privacy, each of which includes the deployment of suitable mitigations and domain-specific quantitative risk assessments (Song et al., 2008).

Modelling threats and analyzing leaks

Even though DP is regarded as the standard of mitigation, we go a step further and do threat modelling to examine the real risk both before and after mitigation. Threat modelling considers the potential methods of assault on an ML system (AMR Research, Inc, Boston, Tech. Rep. 2008). To apply threat [30 modellings], we studied relevant and realistic attacks in a black box environment, such as the tab attack (described below), and we thought about and used innovative attack angles that are extremely pertinent to production models, like the model update attack. We research attacks that attempt more abstract leaking, such as attribute inference, and go beyond the collection of training data. We use those assaults to create privacy metrics once we have built threat models (Acharya et al., 2009).

Attacks using model updates

A Microsoft Research team described a new threat model where a user can access numerous snapshots of a model, such as predictive keyboards, in the paper Assessing Information Leakage from Updates to Natural Language (NL) Models (Mouratidis et al., 2013). They suggested utilizing attempted model attacks to examine data leakage in real-world scenarios where language models are continuously updated by introducing new data, perfecting publicly available which was before language models on personal data, or erasing user data to adhere to privacy laws. The findings demonstrated that access to such samples can reveal the words used to modify the model. Without having to watch it, it is possible to undertake leakage evaluations of text forecasts based on the attack (Gehrke et al., 2010).

Tab Attacks

When an attacker has access to a language model's top-1 predictions and the text auto-completion feature is used, for example, in an email programme, tab attacks can happen. Large language models are known to memorize specific training instances, and recent research has shown that practical attacks can extract verified training examples from GPT-2 (Gregor et al., 2013). A team of Microsoft researchers developed a method to check a language model for training data leakage, which they described in the paper-Training Data Leakage Analysis in Language Models. By utilizing a practical attack, the model builder can determine the degree to which training examples can be pulled from the model using this method. Using this technique, the model owner can check that mitigations are working as planned and decide whether a model is secure to deploy (Goossenaerts et al., 2009).

Poisoning Attacks

Researchers from Microsoft and an allied academic examined the effects of a scenario in which some of the training data were purposefully altered to increase privacy leaks in the study Property Inference from Poisoning. For instance, in a collaborative learning environment when data from several parties or tenants are merged to create a better model and one of the parties is acting dishonestly, this form of a data breach may occur (Gregor et al., 2013). The study provides an example of how such a party could manipulate their data to derive general statistics about the remaining training set. In this instance, a spam classifier is trained using data from multiple parties. Extra care must be taken to guarantee that the data utilized in such joint training settings is reliable because if one of those parties has malicious intentions, it can use the model to determine the average sentiment of the emails in the remaining training set (Goossenaerts et al., 2009).

Computer environments that are private and secure

Customers may be hesitant to upload their data to the cloud at all when dealing with extremely private material. In general, cloud confidential [35 computing makes use of trusted execution environments, which are supported by hardware security guarantees, to make it possible to compute data analytics and ML algorithms on private data while ensuring that cloud administrators, malicious actors who cross the cloud tenancy boundary, and even the cloud provider itself cannot access the data. Multiple clients can work together on private data using this technology without having to have faith in the cloud provider (Eggert et al., 2014). TEEs rely on particular hardware to provide security assurances, however cryptographic secure computing techniques like fully homomorphic encryption (FHE) and secure multi-party computation (SMPC) can process data while still maintaining a high level of encryption. The term "MPC" refers to a collection of cryptographic protocols that enable multiple parties to calculate functions on their shared secret inputs without disclosing anything to each other besides the function's result. FHE is a specific sort of encryption that enables computation to be done directly on encrypted data to prevent anyone else from learning the outcome of the computation. One of the most well-known FHE (Gehrke et al., 2010) libraries, Microsoft SEAL (Mouratidis et al., 2013), was created by a small number of firms.

PPML Tools

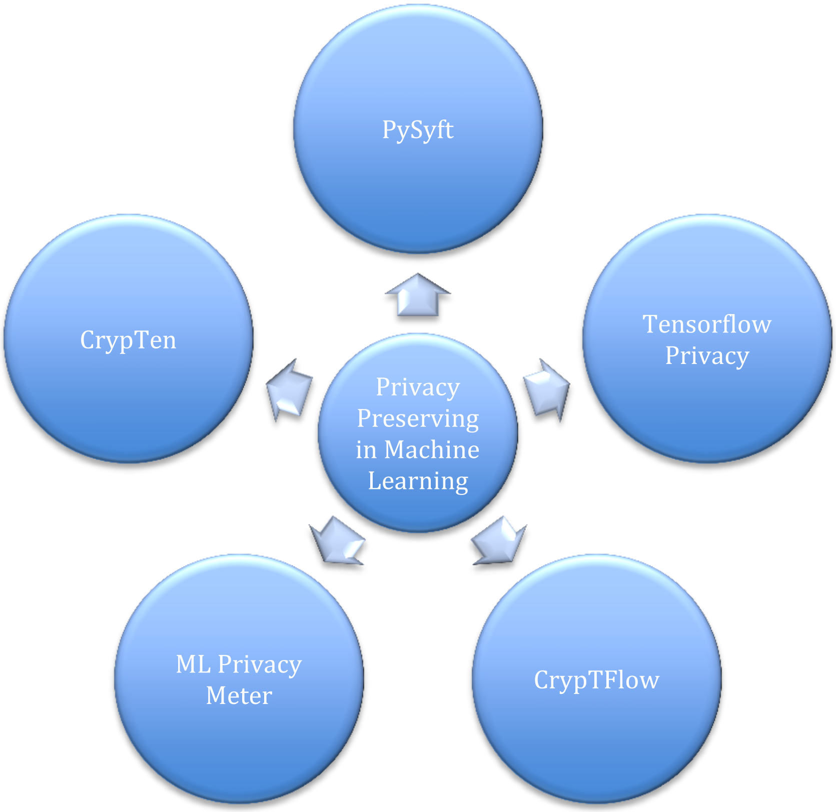

Figure 1: The tools used in Privacy Preserving in Machine Learning (PPML) (AMR Research, Inc, Boston, Tech. Rep. 2008)

1. PySyft

A secure and private machine learning toolbox built on Python is called PySyft. It is a component of the Open Mind programme, which creates technologies and AI frameworks that respect individuals' privacy. The library supports a variety of privacy-preserving methods, such as federated learning, differential privacy, HE, and MPC. PySyft also adds functionality to well-known deep learning frameworks like PyTorch, TensorFlow, and Keras (Gehrke et al., 2010).

2. TensorFlow Privacy

A Python framework called TensorFlow Privacy (TFP) is used to create and train differentially private machine learning models. The library is based on Google's TensorFlow, an open-source machine-learning training framework that disregards privacy issues. An important privacy-preserving ML technique used by the library is to train an ML model using differential private SDG. The differential private mechanism of choice, which can be used to (1) compare ML models in terms of privacy and (2) account for utility loss when selecting one model over another, may also be used to compute the privacy guarantees that it offers (Song et al., 2008).

3. CrypTFlow

CrypTFlow is a framework that offers a method for safely querying ML models using ideas from programming languages and MPC (Acharya et al., 2009).

4. ML Privacy Meter

Both evaluating an ML model's resistance to certain attacks and integrating privacy precautions into the ML process used to build the model are essential. A Python tool called ML Privacy Meter analyses privacy risks in machine learning models using Google's TensorFlow (Sadeghi et al., 2020). The tool can be used to create membership inference attacks under both white-box and black-box adversary models. After that, based on the chosen adversary model, the software may compute the privacy risk ratings. The risk ratings can be used to gauge how precise such assaults on the target model are. The application can also generate privacy reports and show the results (Pitropakis et al., 2019).

5. CrypTen

A machine learning framework for safeguarding privacy is called CrypTen. An open-source machine learning platform called PyTorch is the foundation of the software. With the potential to offer HE (Homomorphic Encryption) in the future, the framework now supports MPC (Bay et al., 2006).

Conclusion

This study discussed data in the aspect of machine learning. By identifying and classifying existing research into three different groups: computer vision, machine learning-assisted personal privacy, and data confidentiality against computer vision threats, review the state-of-the-art methods on this subject and draw numerous findings. The personal deep learning problem has received the most attention. During a review of this type of research work, several propose to employ privacy protection criteria. Due to the difficulty of the data protection aim, the DP style is unable to perform an exhaustive analysis of privacy. Therefore, the issue of how to create new privacy statistics and notes remains unresolved.

References

1. Greener, Joe G., Shaun M. Kandathil, Lewis Moffat, and David T. Jones. "A guide to machine learning for biologists." Nature Reviews Molecular Cell Biology 23, no. 1 (2022): 40-55
2. European Commission. "On Artificial Intelligence-A European Approach to Excellence and Trust." (2020): 1-26.
3. Tabassi, Elham, Kevin J. Burns, Michael Hadjimichael, Andres D. Molina-Markham, and Julian T. Sexton. "A taxonomy and terminology of adversarial machine learning." NIST IR (2019): 1-29.
4. Janiesch, Christian, Patrick Zschech, and Kai Heinrich. "Machine learning and deep learning." Electronic Markets 31, no. 3 (2021): 685-695.
5. Gough, Matthew B., Sérgio F. Santos, Tarek AlSkaif, Mohammad S. Javadi, Rui Castro, and João PS Catalão. "Preserving privacy of smart meter data in a smart grid environment." IEEE Transactions on Industrial Informatics 18, no. 1 (2021): 707-718.
6. McMahon B, Moore E, Ramage D, Hampson E, and Arcas BAY. Communication-efficient learning of deep networks from decentralized data. In Artificial Intelligence and Statistics. 2017;1273-1282.

7. A Goldsteen, G Ezov, R Shmelkin, M Moffie, A Farkash. Data minimization for GDPR Compliance in machine learning models. AI and Ethics. 2022.
8. Nasr M, Shokri R, and Humans A. Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In IEEE Symposium on Security and Privacy. 2019; 1022-1036.
9. Shokri R and Shmatikov V. Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC conference on computer and communications security.2015; 1310-1321.
10. Shokri R, Stronati M, Song C, and Shmatikov V. Membership inference attacks against machine learning models. In Security and Privacy (SP). 2017; 3-18.
11. Song C, Ristenpart T, and Shmatikov V. Machine learning models that remember too much. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017; 587-601.
12. SK Murakonda, R Shokri. Ml privacy meter: Aiding regulatory compliance by quantifying the privacy risks of machine learning. :2007.09339, 2020 .
13. Utilizing machine learning techniques to reveal vat compliance violations in accounting data 2019 IEEE 21st Conference on Business Informatics (CBI).
14. Korba L, Song R, Yee G, Patrick AS, Buffett S, Wang Y, Geng L. Private data management in collaborative environments. In: Luo, Y. (ed.) CDVE. 2007; 4674.
15. Aura T, Kuhn TA, Roe M. Scanning electronic documents for personally identifiable information. In: Proc. of the Workshop on Privacy in the Electronic Society (WPES 2006), Washington, DC, October.2006; 41-49.
16. Agichtein, E., Cucerzan, S.: Predicting accuracy of extracting information from unstructured text collections. In: CIKM 2005. 413-420.
17. Kambhatla, N.Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In: Proc. of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL), Barcelona, Spain. 2004.
18. Miller S, Fox H, Ramshaw L et al. Description of the SIFT system used for MUC-7. In: Proc. of the 7th Message Understanding Conference. 1998.
19. Patel, Chiranjit R. "Luhn Algorithm in 45 nm CMOS Technology for Generation and Validation of Card Numbers." In Inventive Systems and Control, pp. 269-284. Springer, Singapore, 2021.

20. Han H, Giles CL, Manavoglu E, Zha H, Zhang Z, Fox EA.: Automatic document metadata extraction using support vector machines. In: Proceedings of the 2003 Joint Conference on Digital Libraries (JCDL 2003), Houston, Texas. 2003; 37-48.
21. Chang CH, Kayed M, Girgis MR, Shaalan KF. A Survey of Web Information Extraction Systems. IEEE Transactions on Knowledge and Data Engineering. 2006; 1411-1428.
22. Turmo J, Ageno A, Catala N. Adaptive information extraction. ACM Computing Surveys.:2006.
23. Sadeghi, Koosha, Ayan Banerjee, and Sandeep KS Gupta. "A system-driven taxonomy of attacks and defenses in adversarial machine learning." IEEE transactions on emerging topics in computational intelligence 4, no. 4 (2020): 450-467.
24. Duddu, Vasisht. "A survey of adversarial machine learning in cyber warfare." Defence Science Journal 68, no. 4 (2018): 356.
25. Pitropakis, Nikolaos, Emmanouil Panaousis, Thanassis Giannetsos, Eleftherios Anastasiadis, and George Loukas. "A taxonomy and survey of attacks against machine learning." Computer Science Review 34 (2019): 100199.
26. Vorobeychik, Yevgeniy, and Murat Kantarcioglu. "Adversarial machine learning." Synthesis Lectures on Artificial Intelligence and Machine Learning 12, no. 3 (2018): 1-169..
27. Song R, Korba L, Yee G. An Efficient Privacy-Preserving Data Mining Platform. In: The 4th Int. Conf. on Data Mining (DMIN 2008), Las Vegas, Nevada. 2008.
28. McGreevy, “Amr research finds spending on governance, risk management, and compliance will exceed $32b in 2008.”, AMR Research, Inc, Boston, Tech. Rep. 2008.
29. Acharya VV and Richardson M, “Causes of the financial crisis”, Critical review.2009; 195-210.
30. Syed Abdullah N, Sadiq NS, and Indulska M. “Emerging challenges in information systems research for regulatory compliance management”, in Advanced information systems engineering, B. Pernici, Ed., Springer Berlin Heidelberg.2010; 251-265.
31. Eggert M. Compliance Management in Financial Industries: A Model-based Business Process and Reporting Perspective. Springer Science & Business Media. 2014.
32. Massey AK, Otto PN, Hayward LJ, and Antón AI. “Evaluating existing security and privacy requirements for legal compliance”, Requirements engineering. 2010; 119-137.

33. "Mouratidis, Haralambos, Shareeful Islam, Christos Kalloniatis, and Stefanos Gritzalis. ""A framework to support selection of cloud providers based on security and privacy requirements."" Journal of Systems and Software 86, no. 9 (2013): 2276-2293.
34. Gehrke N and Tham R. “VAT Compliance: Einhaltung umsatzsteuerlicher Anforderungen beim Prozess- und Datenmanagement in ERP-Systemen.”, in Mkwi, M. Schumann, L. M. Kolbe, M. H. Breitner, and A. Frerichs, Eds., Universitätsverlag Göttingen.2010; 569-581.
35. 35. Gregor S and Hevner AR. “Positioning and Presenting Design Science Research for Maximum Impact”, Mis quarterly.2013; 337-355.
36. 36. Goossenaerts JBM, Zegers ATM, and Smith JM. “A multilevel model-driven regime for value-added tax compliance in ERP systems”, Computers in industry.2009; 709-727.
37. 37. Bay S, Kumaraswamy K, Anderle MG, Kumar R, and Steier DM.“Large scale detection of irregularities in accounting data”, in Proceedings - ieee international conference on data mining, icdm, IEEE. 2006; 75-86.
38. 38. McGlohon M, Bay S, Anderle MG, Steier DM, and Faloutsos C. “SNARE: A Link Analytic System for Graph Labeling and Risk Detection”, in Proceedings of the 15th acm sigkdd international conference on knowledge discovery and data mining, see. 2009; 1265-1274.
39. 39. Khan RQ, Corney MW, Clark AJ, and Mohay GM. “Transaction mining for fraud detection in ERP Systems”, Industrial engineering and management systems.2010; 141-156.
40. 40. Jans M, Lybaert N, and Vanhoof K. “Internal fraud risk reduction: Results of a data mining case study”, International journal of accounting information systems.2010; 17-41.