ISSN 2616-5775 Vol. 8, Issue 3: 79-86, DOI: 10.25236/AJCIS.2025.080311

Data Risk and Data Compliance Governance for Generative Artificial Intelligence

Zhuoran Li *

School of Economics, Nanjing University of Finance and Economics, Nanjing, Jiangsu, 210023, China
**Corresponding author: Wxfcg2021@126.com*

Abstract: *Generative Artificial Intelligence (Generative AI), being a revolutionary technology, has exhibited varied uses across industries. However, its rapid development has also brought with it high data risks, including leakage of privacy, unauthorized access to data, improper processing of data, abuse of data, infringement of intellectual property rights, and creation of counterfeit information. This paper systematically reviews such risks and explores the current status of generative AI in global and Chinese data compliance regulation. Clarifying existing regulations and highlighting the main challenges, this study presents solutions for enhancing awareness of data privacy protection, enhancing data usage authorization management, enhancing data storage and security functions, and strengthening international cooperation and standardization. The findings suggest the necessity of sound data compliance governance towards the long-term evolution of generative AI and emphasize the need for multidimensional data governance that brings technology, law, and ethics into harmonious alignment.*

Keywords: *Generative AI; data risk; compliance governance; privacy protection; legal regulation*

1. Introduction

Generative AI is a tremendous breakthrough in artificial intelligence that enables machines to produce content that looks very much like human imagination. From language understanding to image generation, generative AI can potentially be used for many applications across different fields of medicine, finance, entertainment,and education. Though its innovative potential is boundless, the very immediate mass growth of generative AI technologies also raises more immediate issues concerning data risk and compliance management. Not only do they threaten the privacy and data security of individuals, but they also have the potential to threaten larger society and economy-wide issues. The understanding and management of generative AI data risks and taking responsibility for them are therefore crucial to sustainable AI development.

In recent years, data risks of generative AI and governance of compliance risks have been researched broadly in academic studies. All literature to date has focused on the protection of privacy, transparency of algorithms, cross-border data transfer, and regulation through the law, etc. Mazurek and Małagocka (2019) state that data protection and privacy are presently challenged by several difficulties in formulating AI [1]. For example, despite having relevant regulations and policies in place, there remains a regulatory gap due to the rapid pace of technological advancements that makes it challenging to cover comprehensively new threats to privacy [2]. In the meantime, the absence of effective practice models for companies to strike a balance between leveraging data to drive AI innovation and ensuring user privacy has resulted in low user trust in companies' data processing practices [3]. Murakonda and Shokri (2020) believed that when utilizing sensitive data to develop machine learning models, organizations need to ensure that the data processed in such systems are properly safeguarded, and emphasize indirect leakage training, suggests ML Privacy Meter, a cutting-edge membership inference attack method, and explains how to assist practitioners in adhering to regulations when implementing it [4]. Brauneck, Schmalhorst, Kazemi Majdabadi, et al. (2023) integrate federated learning with secure multi-party computation and differential privacy combination, believing that the GDPR is required to meet legal data protection obligations in medical research that processes personal data. This blend offers an appealing technical answer for health organizations that are prepared to work together without risking their data. Legally, the combination provides sufficient in-built security mechanisms to satisfy the needs of data protection, and technically, the combination provides a system of security with comparable performance compared to centralized machine learning applications [5]. Christodoulou and Limniotis (2024) have explained the use of machine learning algorithms in an automated decision-making system as introduced

ISSN 2616-5775 Vol. 8, Issue 3: 79-86, DOI: 10.25236/AJCIS.2025.080311

by the data protection issues further elaborating on the use of a technique, i.e., differential datasets. It is stated that it has been illustrated through extensive experiments that some issues arise when one has to establish that the technique is indeed sufficient enough to curb all the threats inflicted on the basic rights of an individual. Whereas the latter can try to achieve up to about 90% using properly tuned parameters of certain algorithms, it's clear that even this figure can be unacceptable for algorithms deciding on the issue-at-hand [6].

Although the recent research is extremely productive, there remains scope for in-depth research in numerous ways. In one sense, there is insufficient deep analysis of the new risk of data with respect to the convergence of new technologies and generative AI, such as the likely security threat resulting from the application of quantum computer technology to the processing of data. On the contrary, scant research has been done on the synergy of cross-domain integral governance methodologies, and how to combine and effectively leverage technological instruments, legal legislation, and moral rules in an organic manner needs further investigation. Meanwhile, in the global coordination of compliance standards, although international organizations have tried to promote relevant work, there are differences in data sovereignty and privacy concepts among nations and regions, and it is required to conduct in-depth research on how to balance these differences and create more universal compliance standards.

The significance of the research for this paper is to systematically integrate the data risk analysis, compliance governance status quo,and strategic suggestions of generative AI, and propose new concepts and insights for the above directions that urgently require in-depth research. First, the risk of generative AI is analyzed from the perspective of multi-dimensional analysis to fully reveal its potential danger and establish a theoretical foundation for further in-depth research on the integration risk of new technologies. Second, the global and domestic compliance governance systems are structured to describe the main challenges in current governance and provide realistic references for optimizing cross-domain integrated governance methods. Finally, the submitted overall governance plan not only adopts technical, legal,and ethical considerations, but also emphasizes international cooperation, and demands providing a feasible way to solve the problem of coordination among international compliance standards and ensuring the sustainable development of generative AI.

2. Analysis of data risk sources for generative artificial intelligence

Generative AI is faced with a series of threats during the process of data collection, processing, and use, which not only threaten individual privacy and data security but also have the potential to cause significant social and economic impacts [7].In the first half of 2024, 1,571 data breaches occurred globally, an increase of 14% over the same period in 2023, and the number of victims was about 1,079 million, close to 6 times that in 2023. The average cost of a data breach worldwide reached a record $4.88 million in 2024, an increase of 10% over 2023. Data breaches are most prominent in industries such as the information and internet industry and the financial industry. The data risks of generative AI as it stands do mainly stem from the following dimensions:

2.1 Privacy leakage risk

Privacy leakage risk is one of the major concerns of generative AI. Generative AI models require lots of data to train, and such data usually contain users' sensitive information. If there are no measures for data protection, such information will be divulged, which will lead to the violation of user privacy. For example, in the healthcare industry, patient health records processed by generative AI models will pose a great risk to patient privacy in case of a leak [8].

2.2 Risk of illegal access

The risk of illegal access is another important consideration. In the process of data collection, there exist some organizations or individuals who gather users' data through illegal methods, i.e., gathering data without users' consent or obtaining data through hacking [9]. These activities not only violate laws and policies but also may lead to data misuse and abuse. For example, some businesses may obtain users' social media data illegally and use it to train generative AI models, thus generating misleading content and influencing public opinion.

ISSN 2616-5775 Vol. 8, Issue 3: 79-86, DOI: 10.25236/AJCIS.2025.080311

2.3 Danger of data mismanagement

Data mishandling risk is yet another significant challenge for generative AI. Human error or technical issues during the data processing session can lead to data distortion or misguided decisions. For example, in the financial industry, when generative AI models are processing large volumes of transaction data, improper data processing can lead to the model generating incorrect prediction results, thus affecting investment decisions.

2.4 Data Misuse and Intellectual Property Risk

Data abuse and intellectual property risks cannot be ignored. The content generated by generative AI models can turn out to be non-informative or even infringe on other individuals' intellectual property rights. For example, some entities may use content generated by generative AI models for commercial advertisements without the consent of the original owners, leading to intellectual property disputes.

2.5 False Information Generation Risk

The risk of false information generation is another essential concern for generative AI [10]. Without cross-references among various sources of information, generative AI models have the ability to invent information or collage data to produce fake content. Such false content might be extensively propagated to deceive the public and even cause social panic. For example, in the news field, generative AI models can generate fake news articles to misguide the public's opinion of events.

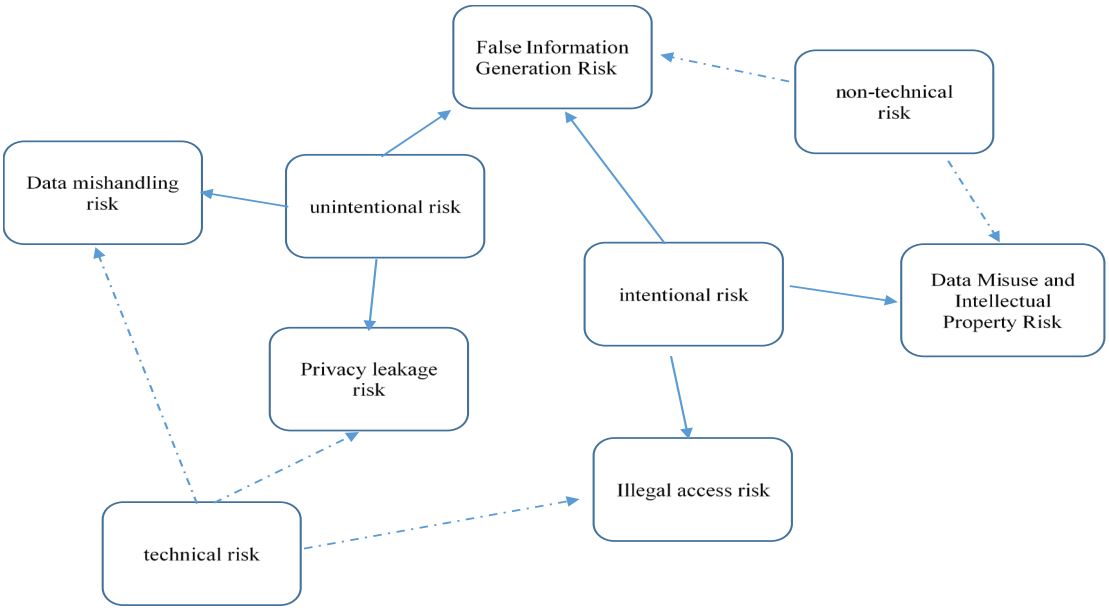

Figure 1 Data risk categories for generative AI

These threats may be categorized into intentional and unintentional threats based on whether the subjective and intended nature of the human being dictates, as shown in Figure 1. Intentional threats involve data abuse intellectual property risk, and criminal access risk. Unintentional threats consist of data mismanagement risk and leakage of privacy risk. Among them, the risk of intentional risk and unintentional risk of generating false information, because the generation of false information belongs to unintentional risk, and the active dissemination of generated false information belongs to intentional risk. Meanwhile, depending on whether the risk is mainly caused by technical problems, the above risks can also be divided into technical risks and non-technical risks. Among them, data mishandling, privacy leakage risk, and illegal access risk are non-technical risks. False information generation risk, data misuse, and intellectual property rights are non-technical risks.

ISSN 2616-5775 Vol. 8, Issue 3: 79-86, DOI: 10.25236/AJCIS.2025.080311

3. Analysis of generative ai data compliance governance regulations

3.1 AI Data Compliance International Governance Regulations

Data governance for compliance in generative AI is mixed globally, with countries and regions following different policies and steps in practice and legislation，as shown in Table 1.

According to the overview of international regulation, areas and nations such as Europe and the United States are pioneers in generative AI data compliance management. For example, the European Union (EU) has set strict data protection and privacy standards in the General Data Protection Regulation (GDPR), requiring companies to obtain users' explicit consent for the processing of personal data and ensure data transparency and traceability. The United States, though, has enacted numerous federal and state legislation and regulations, such as the California Consumer Privacy Act (CCPA), that aims to better protect personal information and give consumers more control over their data. Other countries like Canada and Australia have also enacted similar data protection regulations, forming a relatively comprehensive international compliance governance framework.

In China, generative AI data compliance management has also been prioritized. The Chinese government has implemented various regulations and laws in recent years to regulate the development and use of generative AI. For example, Measures for the Administration of Generative Artificial Intelligence Services clearly delineates the duties and responsibilities of generative AI service providers and requires them to adhere to the rules of lawfulness, legitimacy, and necessity in data collecting, processing, and usage and ensure data security and privacy. In addition, regulations and laws such as the Cybersecurity Law, the Data Security Law, and the Personal Information Protection Law have laid down a solid foundation of laws to govern data compliance for generative AI. The regulations not only require enterprises to establish a sound data protection system but also clearly indicate legal liability for violation, thus having a strong assurance for the healthy development of generative AI.

Table 1 Generative AI Data Compliance Governance Regulatory Categories

Category	Representative Regulations	Characteristics
Integrated Comprehensive Regulatory Class	*Interim Measures for the Administration of Generative Artificial Intelligence Services in China*, *EU Artificial Intelligence Bill*	Multi-dimensional specification of many aspects of generative AI. Emphasize collaboration
Data Privacy Protection	*EU General Data Protection Regulation (GDPR)*, *China's Personal Information Protection Act*	Strict personal data protection. Comprehensive scope of data protection: clear data subject rights
Innovation and Development Promotion	*U.S. National Artificial Intelligence Initiative Act, Japan AI Strategy 2025*	Promote the development of technical standards, focusing on application and development in key areas such as healthcare, transportation, and manufacturing; balance innovation and regulation
Ethics and Principle Oriented	*Singapore AI Governance Framework*	Provide an ethical and value-oriented approach to the development of generative artificial intelligence, integrating multiple factors such as technology, impact on society, and human beings

3.2 Challenges to International AI Data Compliance Regime

Despite the progress in global and Chinese data compliance regulation of generative AI, the latter remains plagued by many challenges.

Firstly, the war between rapid technological advancement and lag in law is becoming increasingly relevant. The rate of iteration of generation AI technology has been much faster than the evolution and renewal of laws and regulations, making it difficult for the existing legal framework to fully cover the risks and challenges of the new technology. For example, generative AI models will encounter complex intellectual property issues when producing content, and existing intellectual property law infrastructure becomes overburdened in managing such issues.

ISSN 2616-5775 Vol. 8, Issue 3: 79-86, DOI: 10.25236/AJCIS.2025.080311

Second, the regulatory challenge due to algorithmic black boxes is also a major one. The decision-making process in generative AI models will not be transparent and interpretable, and it is difficult for regulators to monitor and assess their compliance effectively. For example, in banking, generative AI models can generate advanced investment recommendations, and the process of generating such recommendations is difficult for regulators to understand and examine, thus increasing the complexity and risk of regulation.

Besides, cross-border data flow-induced compliance issues cannot be overlooked. Training and application of generative AI models typically involve cross-national and territorial data flows, and data protection laws are different in countries and regions, posing a huge challenge to the compliance governance of companies. For example, the EU's GDPR requires data to have strong protection levels when crossing borders, while the protection level of data may be very low elsewhere in countries and regions, thus posing stringent compliance requirements for businesses when conducting cross-border data transfers.

4. Data compliance governance enterprise practices and issues explored

4.1 Data Compliance Governance International Corporate Practice Cases

As for compliance governance of generative AI, much useful and successful experience has been gained domestically and abroad and can be a useful reference for the development of the industry.

With the advent of the General Data Protection Regulation (GDPR) in the EU, the market environment has undergone drastic changes. Many companies have quickly realized the strategic value of protection of data privacy and management of compliance, and are engaging proactively to upgrade and streamline related initiatives. Large technology companies such as Google and Microsoft, for example, have demonstrated exceptional vision and execution in tackling GDPR compliance requirements. In the construction of data protection mechanisms, they have invested a lot of resources in technical research and development and process optimization. With the data collection source, they have designed a thoughtful user notice procedure, clearly stating critical information such as the purpose of data, retention period, and likely sharing recipients, thus enabling users to make independent and clear authorization choices based on full awareness. In data storage, the most recent encryption algorithms are used, such as the widespread application of the Advanced Encryption Standard (AES), which provides multi-level encryption of the user data, making it extremely difficult to crack and utilize the data even if stolen illegally during storage. At the same time, the open data processing mechanism has been built meticulously, taking advantage of the block chain technology's traceability feature to trace in detail each data call, analysis, processing and other operation, and opening up the query rights for users. In the process, not only does it meet the strict compliance requirements of the GDPR in an exhaustive way, but more importantly, raises users' confidence in data protection levels all around the world to a much higher level. The above-mentioned salutary compliance culture also assisted the company in building a good reputation, enhancing its competitive edge in the marketplace, and encouraged more privacy-conscious users to utilize its products and services.

In China, ever since the official enforcement of the Measures for the Administration of Generative Artificial Intelligence Services, domestic companies have also actively responded and done their best to facilitate the enforcement of compliance governance. Baidu and Tencent, among others, have pioneered in introducing advanced data encryption technologies because of their strong technical R&D capability. For example, Baidu has also forayed into the use of quantum encryption technology on its AI data processing platform, taking quantum mechanics concepts to the next level to accomplish completely secure data storage and transmission, thus effectively eradicating data theft or interference threat. At the same time, Tencent has developed highly rigorous access controls with granular and precise segregation of privilege in association with users' roles, responsibilities, and business needs. Utilizing technology such as multi-factor authentication and dynamic access tokens, Tencent only permits approved individuals to access specific data resources within a limited space and time. Not only do these actions effectively reduce the risk of data leakage and abuse, but they also set a benchmark for the industry as a whole, and urge other companies to step onto the road of compliance governance, and promote energetically the healthy and orderly growth of the domestic generative AI industry.

4.2 Data Compliance Governance International Business Practice Issues

Although the above success stories are good empirical paradigms for generative AI compliance

ISSN 2616-5775 Vol. 8, Issue 3: 79-86, DOI: 10.25236/AJCIS.2025.080311

governance, there are still a series of complicated and challenging challenges in real application scenarios. At the level of enterprises themselves, some enterprises have exposed a series of problems in AI data compliance governance. Due to insufficient focus on compliance governance and an absence of well-planned strategic management, there are certain businesses that are far too laid back in data gathering and fail to thoroughly fulfill their responsibility to inform users, resulting in a huge amount of personal data being gathered without the knowledge of the users, which violates the right to know and autonomous choice of the users.

As far as data storage is concerned, some firms are limited by costs, fail to update and upgrade data storage software and hardware in a timely manner, and employ a less secure storage architecture, which is very easy to become the target of hacker attacks, greatly increasing the likelihood of data leakage. At the same time, there are some companies without sound monitoring and auditing systems in the data processing process, the data processing process is in randomness and disorder, unable to ensure the accuracy and integrity of the data, and even may be induced by improper data processing means caused by data misuse problems.

Other than that, there are no serious reserves of data governance compliance skills within certain companies, nor are there any composite professionals with knowledge on AI technology coupled with expertise under legislation and regulatory issues. This leads to interpretation and application bias for relevant compliance policies, cannot develop an effective compliance governance program, and are prone to fall into passive response to regulatory audits and inspections. These problems are inherent within the enterprises themselves, heavily hindering effective promotion of generative AI data compliance governance, and should be specifically emphasized and resolved as a matter of urgency.

5. Data compliance governance strategies for generative artificial intelligence

5.1 Enterprise Strategy Level

In order to effectively deal with the various risks faced by generative AI in data collection, processing, and use, it is necessary to build a comprehensive data compliance system.

5.1.1 Strengthen the awareness of data privacy protection

Strengthening the awareness of data privacy protection is imperative. Companies should raise the awareness of data privacy protection among employees and users with ongoing training and promotion so that the data protection principles are followed at all stages of data processing. The need to improve data collection and processing practices is critical. Companies should formulate detailed data collection and processing requirements to define the scope, purpose and manner of data collection and guarantee data legitimacy and legality. At the same time, an emergency response mechanism for data leakage is required. Enterprises should formulate emergency response plans and conduct regular drills so that they can respond in a timely fashion in the event of data leakage and minimize losses.

5.1.2 Strengthen data use authorization management

Strengthening data use authorization management is another important step. Classification management of users' data is needed. Enterprises should categorize data in accordance with its sensitivity and usage scenarios and formulate corresponding management measures. For example, for sensitive data, stronger protection should be adopted, such as encrypted storage and access control. Meanwhile, establishing a sound authorization record system is also crucial. Enterprises should record the authorization of every use of data to ensure that every use of data is logged. Additionally, strengthening data usage audits and inspections is also necessary. Organizations must conduct periodic data use audits to ensure data is used in a legal manner and that violations are detected and dealt with in good time.

5.1.3 Enhance data storage and security capabilities

Enhancing data storage and security capabilities is an important part of data compliance governance. The use of advanced security technologies is crucial. Companies should use the latest data encryption, firewall, and intrusion detection technologies to secure data during storage and transmission. And it is also necessary to have a data backup and recovery process. Companies should backup important data regularly and should have a solid data recovery plan for quick recovery in case of data loss or corruption. In addition, enterprises need to implement a real-time monitoring system for overall monitoring of data storage and security to enforce data security and compliance.

ISSN 2616-5775 Vol. 8, Issue 3: 79-86, DOI: 10.25236/AJCIS.2025.080311

5.2 Legal and Regulatory Aspects

5.2.1 Specialized Generative AI Data Compliance Legislation Design

With the pace of generative AI development accelerating ever more rapidly, existing data compliance provisions scattered in different laws and regulations can hardly meet the industry's demand. It is recommended that the government creates a data compliance law for generative AI specifically to regulate systematically and comprehensively the data life cycle management in all respects. The law should specify the generative AI enterprise's right to data and responsibilities, expand on particular requirements and standards for compliance with data, increase the seriousness of penalties for data offense, and increase the cost of offending. Meanwhile, the law must look to the future, have the capability of keeping up with the round-the-clock technology upgrade and innovation, and have the ability to develop long-term stable legal protections for industry growth.

5.2.2 Promote synergistic technical standard and policy making

Policy made and technical standards must be interlinked closely so that synergistic effects can be generated. In the one case, policy makers must improve communication and coordination with industrial associations and technical professionals in order to fully grasp the trend and characteristics of generative AI technology development, thus enabling policy provisions to capture the needs of technological development in an effective way. On the contrary, industry associations and standardization organizations are called upon to create relevant technical standards, e.g., data security technical standards, data quality standards, data labeling standards, etc., in an effort to provide the relevant technical guidelines for enterprises' data compliance practice. The policy should explicitly require enterprises to comply with the relevant technical standards and consider technical standards as an important facilitation of policy implementation.

5.2.3 Promote international cooperation and policy harmonization

International cooperation and policy harmonization are international measures to address generative AI data risks. Firstly, international cooperation is needed for the global community to set data protection and compliance governance standards for generative AI. Through international cooperation, countries are able to share experiences and best practices and develop harmonized compliance governance standards in order to facilitate the healthy development of generative AI globally. Second, promoting the sound development of technology is also necessary. The international community should increase technical exchanges and cooperation to promote the innovation and application of generative AI technology, ensuring the safety and compliance of the technology. Besides that, a mechanism for transnational data flows' compliance must be established. By bilateral or multilateral agreements, states should establish compliance standards for cross-border data flows to ensure the security and legitimacy of data as it crosses borders.

6. Conclusion

This paper thoroughly analyzes the various risks generative AI is subject to in data collection, processing, and use, and sorts through the current Chinese and international regulatory regime and experience in generative AI data compliance regulation. By comparative analysis, it reveals the most important challenges that generative AI data compliance governance is facing currently, and synergistically propose corresponding measures at the enterprise level and the regulatory level. At the enterprise level, corresponding measures are suggested to improve data privacy protection awareness, make the management of authorizing data use more effectively, and enhance the capability of storing and protecting data. At the institutional level, countermeasures are put forward from the angles of establishing specialized generative AI data compliance legislation, encouraging policy and technical standard synergistic development, enhancing international policy cooperation and coordination, and raising global public awareness. The research indicates that good data compliance governance is essential for the normal development of generative AI. In the future, it is required to improve further the technology, law, and ethics synergistic governance that is multidimensional to address the complex data challenges and risks created by generative AI. Data compliance governance of generative AI is given theoretical foundation and practical guidance by research in this paper, which possesses high theoretical and practical significance. However, due to the limited quantity of data information, the research remains insufficient in quantitative analysis. In the sequel, research will continue to consolidate the assembly of relevant data information, validate the relevant concepts from empirical and theoretical analysis perspectives, and provide more innovative recommendations.

ISSN 2616-5775 Vol. 8, Issue 3: 79-86, DOI: 10.25236/AJCIS.2025.080311

References

[1] Grzegorz Mazurek, & Karolina Małagocka (2019). Perception of privacy and data protection in the context of the development of artificial intelligence. Journal of Management Analytics, 6 (4), 344-364. https://doi.org/10.1080/23270012.2019.1671243)

[2] Abel Monfort,Mariano Méndez Suárez & Nuria Villagra.(2025).Artificial intelligence misconduct and ESG risk ratings.Review of Managerial Science. https://doi.org/10.1007/s11846-025-00850-9

[3] Mohammad Gouse Galety,Jimbo Henri Claver,A. V. Sriharsha,Narasimha Rao Vajjhala & Arul Kumar Natarajan.(2024).Data Analytics and AI for Quantitative Risk Assessment and Financial Computation. https://doi:10.4018/979-8-3693-6215-0.

[4] Sasi Kumar Murakonda, & Reza Shokri (2020). ML Privacy Meter: Aiding Regulatory Compliance by Quantifying the Privacy Risks of Machine Learning.. arXiv (Cornell University).

[5] Alissa Brauneck,Louisa Schmalhorst,Mohammad Mahdi Kazemi Majdabadi,Mohammad Bakhtiari, Uwe Völker,Jan Baumbach,Linda Baumbach, & Gabriele Buchholtz (2023). Federated Machine Learning, Privacy-Enhancing Technologies, and Data Protection Laws in Medical Research: Scoping Review. Journal of Medical Internet Research, 25 (0), e41588. https://doi.org/10.2196/41588

[6] Paraskevi Christodoulou, & Konstantinos Limniotis (2024). Data Protection Issues in Automated Decision-Making Systems Based on Machine Learning: Research Challenges. Network, 4 (1), 91-113. https://doi.org/10.3390/network4010005.

[7] Amit Arora,Michael Barrett,Euisin Lee,Eivor Oborn, & Karl Prince (2023). Risk and the future of AI: Algorithmic bias, data colonialism, and marginalization. Information and Organization, 33 (3), 100478. https://doi.org/10.1016/j.infoandorg.2023.100478

[8] Kashif Naseer Qureshi,Hanaa Nafea, & Pyoung Won Kim (2024). Advancing healthcare systems: A tri-tier architecture by using data communication, AI data generative and regulation and compliance standards. Expert Systems. https://doi.org/10.1111/exsy.13742

[9] Niklas Kruse & Julius Schöning.(2024).Legal conform data sets for yard tractors and robots: AI-based law compliance check on the right to one's image. Computers and Electronics in Agriculture. https://doi.org/10.1016/j.compag.2024.109106.

[10] Amanda Heidt.(2024).Intellectual property and data privacy: the hidden risks of AI..Nature. https://doi.org/10.1038/d41586-024-02838-z