NextGen Scientific Publication
American International Journal of Computer Science and Technology
Volume 1 Issue 3, Pg. No. 12-22, AIJCST-V1I3P102, 2019
https://doi.org/10.63282/3117-5481/AIJCST-V1I3P102

Original Article

Enabling End-to-End User Data Privacy Compliance for GDPR/DMA Using Machine Learning

***Deepak Venkateshappa**
Staff Data Engineer, San Jose, California, USA.

Abstract:

The rapid growth of digital platforms and cloud systems has increased personal data collection. Regulations like GDPR and DMA enforce strict data privacy and transparency rules. Organizations face challenges in achieving end-to-end compliance across complex systems. Traditional manual compliance methods are not scalable or adaptable. There is a need for automated and intelligent compliance solutions. The paper proposes an ML-based workflow for privacy regulation compliance. It includes data discovery, classification, consent checking, and anomaly detection. The system identifies sensitive data and monitors access patterns. It uses supervised, unsupervised, and reinforcement learning techniques. Explainable AI ensures transparency, accountability, and auditability. A layered model includes data ingestion, risk scoring, and governance reporting. NLP and deep learning enable automated data labeling. Clustering detects unusual access patterns and potential violations. Federated learning ensures privacy during distributed model training. Results show improved accuracy, faster compliance, and reduced privacy risks.

Keywords:

GDPR, DMA, Data Privacy, Machine Learning, Compliance Automation, Explainable AI, Privacy Risk Scoring, Data Governance, Federated Learning, Anomaly Detection.

Article History:

Received: 18.03.2019

Revised: 08.04.2019

Accepted: 26.05.2019

Published: 18.06.2019

1. Introduction

1.1. Background

The growing size of the digital platform, cloud computing, artificial intelligence and networked services have turned the global economy into a highly data-driven ecosystem. [1,2] Companies in any business sector are now capturing, processing, and trading with large amounts of personal and behavioral data, consisting of distinctive identifiers, movement history, browsing histories, biological markers, monetary dealings, and user-created content. Although this data-based revolution has made innovation, personalization, and operational efficiency possible, it has also increased the protectionism associated with privacy breaches, mass surveillance, algorithmic profiling, and massive data breaches. To counter such risks, regulatory offices have come up with detailed legal documents like the General Data Protection Regulation (GDPR) and the Digital Markets Act (DMA) that enforce strict conditions on legal processing of data, transparency, restrictions in the purpose, management of user consent, accountability and fair digital market practices. Although the regulatory goals are clear, it is still a daunting task to convert theoretical requirements into real world implementation in the complicated enterprise structures. The current status of the information systems can be described as decentralized microservices systems, hybrid systems, multi-cloud systems, edge computing systems, mobil ecosystems, and integrating with third parties on a large scale. Information is constantly transmitted between departments inside the organization and between the organization and its suppliers, analytics pipelines, cross-border networks, and frequently without a centralized control system.

Old forms of compliance, which are mostly relying on fixed rule engines, periodic audits, paper-based documentation, and event based incident reporting, find it difficult to maintain reliance on the dynamism and faster speed and disparity of modern

data environments. These methods often result in a large number of false alarms, insufficient flexibility, and a slow rate of identifying noncompliance. In that regard, machine learning comes out as a change-making facilitator of smart privacy regulation. Advanced models have the capability of automatically grouping sensitive data, tracking user actions, establishing abnormal patterns of access, forecasting regulatory risks, and always responding to the changing circumstances of operation. Organizations can replace reactive and checklist-based compliance with proactive, real time governance by incorporating adaptive learning advanced into compliance operating systems. This kind of smart architecture allows not only to improve compliance with regulations but also to work more effectively as an organization and gain the confidence of stakeholders in an ever more controlled digital environment.

1.2. Importance of Enabling End-to-End User Data Privacy Compliance

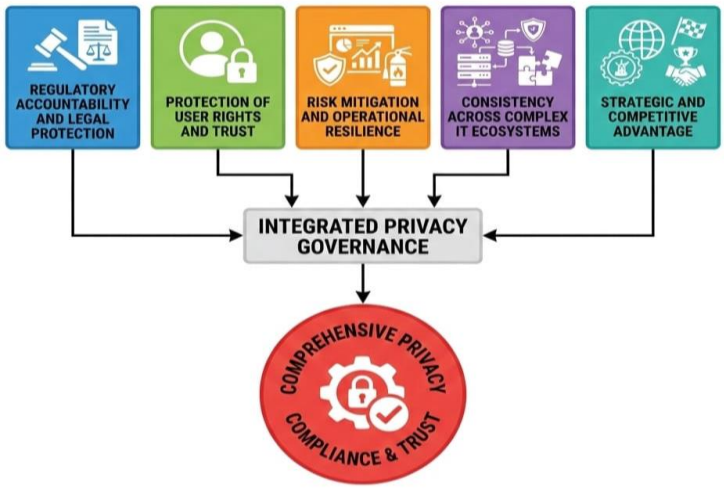

Figure 1. Importance of Enabling End-to-End User Data Privacy Compliance

1.2.1. Regulatory Accountability and Legal Protection

The compliance of the end-to-end user data privacy ensures the compliance with the overall regulatory frameworks including GDPR and DMA. [3,4] These rules carry an extremely important requirement on organizations to be in a transparent state, seeking proper consent, processing lawfully, and protection of data subject rights. Failure to comply may attract hefty fines, legal sanctions, blocking of operations and indirect costs of image. End-to-end compliance strategy makes regulatory principles to be integrated across the whole data lifecycle, including as of data collection and storage, processing, sharing, and deletion, making regulatory accountability mechanisms a top priority and reducing the risk of legal liability.

1.2.2. Protection of User Rights and Trust

Consumers in the contemporary strategic landscape have become more conscious of their privacy rights and demand management of their personal data in a responsible manner by an organization. Basic rights including access to data, amendment, deletion, mobility and protest against processing are end-to-end compliance protections. Organizations that apply open data policy and respect users choices throughout all the systems develop trust and relationships with customers in the long-term perspective. Credibility between the digital market has already become a competitive standout and robust privacy policies enforce the credibility of the brand, as well as the confidence of users.

1.2.3. Risk Mitigation and Operational Resilience

The full privacy compliance mitigates the risk of the leakage, unauthorized data access, and abuse of personal data. Always tracking the flow of data, controlling access, and verifying consent can help organizations not only predict possible risks and eliminate them early but also prevent them from turning into regulatory incidents. End-to-end compliance also increases operational resiliency due to the manner in which privacy controls are built into system architectures instead of being an external oversight capability. This combined strategy also means that compliance is dynamically updated to changing technologies and threat environments.

1.2.4. Consistency across Complex IT Ecosystems

The enterprise information technology space can be highly disaggregated between the cloud environment, third-party services, and mobile apps, as well as an old system. In the absence of an end-to-end compliance structure inconsistencies in how they enforce their policies and data handling practice may emerge. A single compliance architecture means uniform

implementation of privacy controls, standardized classification systems, and single visibility of distributed infrastructures. This integrated management enhances effectiveness of governance and narrows down gaps in compliance.

1.2.5. Strategic and Competitive Advantage

Outside regulatory requirement, end-to-end privacy compliance has strategic advantages. An organization with good privacy governance is also able to form partnerships, enter controlled markets, and distinguish itself within competitive digital markets. Privacy via design and default, similar to technological innovation, will provide the data-centered economy with the opportunity to grow sustainably as it gains a sense of responsibility and accountability.

1.3. Compliance for GDPR/DMA Using Machine Learning

The regulations of the GDPR and the DMA processes will demand the constant monitoring, transparency, responsibility, and severe control of the process of collecting, processing, storing, and sharing the user data. The conventional compliance tools, based on fixed rules, manual auditing, and periodic tests, do not work in the dynamic digital ecosystems where information is transferred between distributed systems in real time. The approach to solve these issues with the help of machine learning is to implement intelligent automation in the compliance processes and make them scalable and adaptive. Machine learning systems can also automatically detect personally identifiable information and susceptible features on structured and unstructured data through high-quality data classification models in the run of which regulated data is adequately labelled and secured. The techniques of natural language processing also improve the skills in the interpretation of privacy policies, consent records, and contractual agreements, which aligns the operational practices to those of law. [5] Besides that, anomaly detection algorithms facilitate constant observation of behavior by detecting the unusual access patterns, unauthorized transfer of data, or suspicious processing activities, which might lead to a violation.

Predictive risk modeling facilitates upstream governance by making a risk estimate of non-compliance, depending on risk context like data sensitivity, cross border transfers and third party integrations. Such potentials enable the organizations to go beyond incident management which is reactive to the prevention of compliance. Machine learning further enhances practices related to DMA adherence which are running platform behaviors, sharing data and gatekeeper requirements; such actions presuppose fair digital competition and transparency. Together with explainable AI approaches, ML-based compliance systems will be able to offer understandable reasons behind automation decisions, which is essential in regulatory settings requiring auditing and accountability. The machine learning models can provide flexible policy administration and scalable governance by constantly learning about new models of data and new changes to regulations. Finally, using machine learning to comply with GDPR and DMA is a privacy management approach that goes beyond being intelligent, real-time, and enterprise-wide, conforming to the current legal standards.

2. Literature Survey

2.1. Regulatory Compliance Automation

Early regulatory compliance engines were mostly rule-based engines that had the set policies and requirement regulations to enforce. [6] The systems were based on fixed sets of rules, the matching of keywords, and the structured workflows that identified the violation. Although good in clear-cut situations, they could not cope with flexibility in dynamic regulatory situations where laws and interpretations have the tendency to change. Consequently, they tended to produce high false-positive rate in such systems and therefore the validation process involved a lot of human intervention. Before 2021, the most common area of research involved how to enhance automation with metadata tagging, policy encoding as standardizations, and control mapping in its structure. Such methods enhanced efficiency and were still not useful in processing unstructured data and contextual regulatory nuances.

2.2. Machine Learning in Privacy Governance

Machine learning has also altered the privacy governance greatly since it provides smart classification of data, tracking, and predictive risk evaluation. [7] Algorithms based on Natural Language Processing (NLP) have been applied several times to detect personal and sensitive data on structured and unstructured information, better than using keywords-based algorithms. Non-compliance likelihood has been estimated using the probabilistic models and regulatory exposure is estimated. Also, clustering and anomaly detection algorithms have been used to facilitate behavioral monitoring by detecting abnormal patterns of access to data or anomalous processing behavior. The model of deep learning has even improved performance in classification, when it comes to complex data. Nevertheless, those enhancements also presented the issues of the model interpretability, accountability and regulatory transparency, particularly in highly regulated sectors.

2.3. Privacy Risk Quantification Models

Privacy risk quantification models seek to systematically assess compliance exposure using a combination of multiple risk factors into a systematic scoring system. [8] The variables that are usually regarded in these models are data sensitivity, context of processing, frequency of access, the scope of regulations, and previous incidences. Through the relative prioritization of the various factors and the probability of the regulatory violation, organizations are able to create composite risk scores that determine mitigation policies. In spite of the fact that such models offer quantifiable decision-making guidelines, most existing frameworks are based on fixed weighting models and periodic evaluations. This constrains their capacity to be flexible in order to adapt to dynamic demands of evolving regulations, organizational developments, or development of threat environment. Therefore, continuous recalibration mechanisms are one of the areas that need additional research.

2.4. Explainable AI in Compliance

Explainable Artificial Intelligence (XAI) has become an essential element in ML-driven compliance systems, especially in regulatory areas that require transparency and auditability. [9] Models Model agnostic explanation methods and feature attribution systems are techniques that allow stakeholders to have an idea of why a model made a given decision on compliance or risk rating. These strategies promote trust, regulatory audits and internal governance reviews. XAI will close the divide between the technical output of ML and legal or policy-inspired reasoning by giving interpretable insights into model behavior. Nevertheless, it is difficult to strike a balance between the performance of the model and its interpretability, especially in versatile deep learning models.

2.5. Research Gaps

Regardless of the development in automation, machine learning, and explainability, there are still research gaps. Little has been done in terms of comprehensive end to end compliance architecture that combines data ingestion, risk modeling, risk monitoring and risk reporting into a single adaptive architecture. Many of the existing systems do not yet comply with the future regulatory standards, including the Digital Markets Act (DMA). Also, federated learning, where sensitive data are not concentrated but model training is performed jointly has not appeared in privacy compliance solutions. Lastly, compliance metrics have not been standardized and unified enough to offer uniform cross-organizational benchmarking. The need to address these gaps is necessary to create scalable, adaptive, and regulator ready privacy governance systems.

3. Methodology

3.1. System Architecture

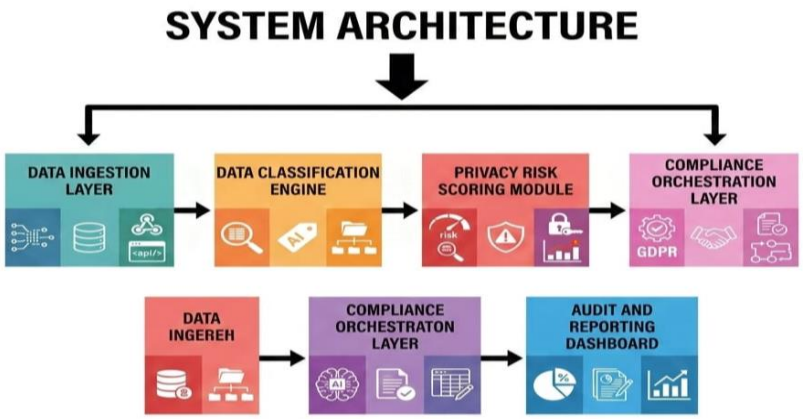

Figure 2. System Architecture

3.1.1. Data Ingestion Layer

The Data Ingestion Layer is the system entry point and it has the duty of collecting data of various sources including internal and external sources. [10,11] Such sources can be databases, cloud storage systems, APIs, enterprise applications and log management platforms. The layer also supports the structured and unstructured data formats and as such, documents, emails, transaction records and user activity logs are compatible. It does basic preprocessing operations like normalization of data, validation, deduplication and metadata. This layer achieves this by standardizing incoming streams of data, so that those components that it serves feed on in downstream are of a consistent high quality input to do more analysis.

3.1.2. Data Classification Engine

The Data Classification Engine is used to examine the data received and classify sensitive and personal data. The engine identifies personally identifiable information (PII), financial data, health records and other regulated information using machine learning and natural language processing methods. It gives the classification labels that are defined according to the regulatory requirements and organizational policies. The engine keeps getting better and better with retraining of the models and with feedback loops which increase the accuracy of the detection over time. Adequate categorization allows implementing privacy settings specifically and minimize the chances of wrongful treatment of data.

3.1.3. Privacy Risk Scoring Module

Privacy Risk Scoring Module compares the degree of compliance with respect to the known data assets and process operations. It takes into account several risk factors including sensitivity of data, its purpose of processing, storage location, regularly accessed, shared by third party and past occurrences. When all these factors are combined, the module produces a composite score of the risk, which is used to indicate the probability and potential consequences of non-compliance. The scoring mechanism facilitates dynamic recalibration, which enables dynamically updating in case regulatory conditions or operational situations change. This module will help decision-makers develop priorities in mitigation strategies and make sure that compliance resources are well-allocated.

3.1.4. Compliance orchestration Layer

Compliance Orchestration Layer converts risk knowledge into action control measures. It automatically implements policies, activates alerts relating to risky actions, and puts remedial processes into operation when a non-compliance occurrence notices are identified. It is a layer that works with the enterprise governance systems to impose either access restrictions, data minimisation rules, encryption protocols, or retention policies. It also provides the correspondence to the applicable regulations, including GDPR or DMA, by mapping system activities to regulatory requirements. This layer will minimize the number of people who will handle the coordination manually and increase the responsiveness of regulations.

3.1.5. Audit and Reporting Dashboard

The profile of compliance status and system performance is monitored through a centralized interface provided by the Audit and Reporting Dashboard. It represents visual risk scores, classification results, trend of incidents, and regulatory mappings in a format that is easy to read. Audit trails through the dashboard assist in preserving a record of every decision, change in policies, and remediation activity. The flexibility in reporting enables organizations to produce compliance reports to the regulators, internal auditors and the executive stakeholders. This component helps in enhancing accountability and facilitates ongoing governance progress, as it ensures transparency and traceability.

3.2. Machine Learning Models

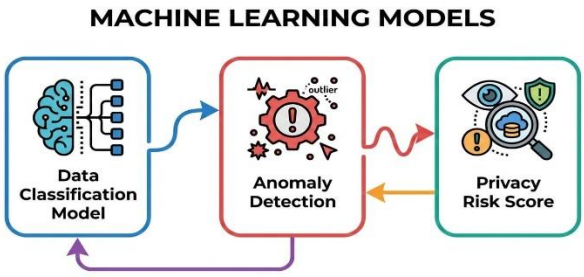

Figure 3. Machine Learning Models

3.2.1. Data Classification Model

Data Classification Model takes a hybrid strategy of a combination of Natural Language Processing (NLP) and a Deep Neural Network architecture. [12, 13] This model uses input data and converts it into numerical feature vector in which textual patterns, semantic meaning, and contextual indicators are evaluated. These characteristics are subsequently fed into several neuron network layers where the weighted transformations and non-linear activation functions allow the model to acquire the complicated relationship in the data. A non-linear intermediary activation mechanism is used, which enables the model to pickup the subtle differences between sensitive and non-sensitive information. Lastly, a sigmoid activation function generates the probability score of whether or not the information falls into a specific privacy classification or not. This architecture boosts the accuracy of classification but it is flexible with different type of data.

3.2.2. Anomaly Detection

To detect anomalies, the system uses the Isolation Forest algorithm in order to detect unusual or potentially threatening behavior of accessing data. This procedure works by randomly dividing the data points and determining the ease with which a given observation can be segregated out of the remaining data. Partitions that make fewer observations to isolate are observed to be more anomalous since they are much different when compared to normal behavior patterns. The algorithm is used to produce a score of anomaly by the average path length needed to isolate a data point relative to what the data should have done. As part of privacy governance, this method can be used to identify anomalies in the access frequency, abnormal user activity, or suspicious processing behavior that can signal that the compliance and insider threat.

3.2.3. Privacy Risk Score

Privacy Risk Score (PRS) is a composite measure which aims at measuring overall exposure of privacy. It is estimated using a weighted average of the three main elements which are the score of data sensitivity, the measure of access frequency anomaly, and the consent risk indicator. The data sensitivity score demonstrates the level of criticality or regulation on the information e.g. individual identifiers or financial records. The access anomaly component is used to quantify abnormalities of normal behavior during usage, emphasizing on abnormal data interactions. The validity risk assesses the adequacy of user approval, and its consistency with the legal provisions and regulations. Time parameters enable weighting of each factor followed by giving importance to them by the organizations, as the model can be adapted by the organizations depending on regulatory priorities or risk tolerance of the company. This systematic method of scoring allows prioritization of risks systematically and the ability to make informed choices on compliance.

3.3. Federated learning framework

Federated Learning Framework is such a model that it helps to train a model and collaboratively use the results of the model learning among multiple organizations or distributed data sources without using a centralized source of data collection. [14,15] Individual participating nodes in this process, which can be a department, subsidiary, and partner organization, will be trained by training a localized version of a machine learning model on its own private data. When sharing raw data, it is possible to risk violating privacy policies or organizational policies, so each node proceeds to compute new model parameters directly. Such local updates are then sent to a central aggregation server. Computing the weighted average of the parameters received at all the nodes participating in the global model is used to update the model. The strengths of each local model are proportional to the size of the dataset utilized at that node, so that nodes with bigger and more illustrative dataset make proportional contribution to the overall model. Mathematically, the new global model at the next training round is computed as the sum of the local model parameters of all the participating nodes, with each parameter being multiplied by the fraction of the dataset size of the node to its total dataset size summed over all the nodes. This additive combination method proceeds until the model is reduced to some optimal or satisfactory level of performance. Having this raw data remain localized and furthermore, only transmitting model updates, federated learning minimizes the risk of data exposure and regulatory non-compliance. This model is best applicable in privacy governance systems applied in more than a single jurisdiction where the data transfer is highly restricted. Moreover, it allows the provision of scalability and flexibility because new participants may enter the training process without impacting the entire architecture. Nevertheless, it might need secure aggregation mechanisms and techniques of differential privacy, to further protect model updates against such attacks of inference.

3.4. Flow of Compliance Enforcement

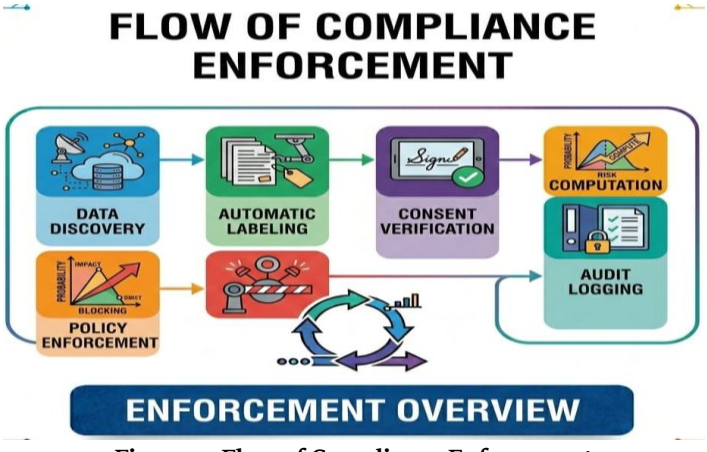

Figure 4. Flow of Compliance Enforcement

3.4.1. Data Discovery

The first on the compliance enforcement process is the data discovery, during which the system detects and identifies data assets in the organization. [16,17] It entails going through the databases, cloud repository, enterprise applications, terminals, and communication infrastructure to identify both systematized and unsystematized information. State-of-the-art tools of discovery are based on the idea of pattern recognition and metadata analysis to identify sensitive data, including personal identifiers, finances, or health-related data. This is aimed at developing a detailed data asset inventory, such as location and format of data, data ownership as well as processing context. Good and efficient data discovery forms visibility that is crucial in ensuring regulatory compliance and hidden risks are minimized.

3.4.2. Automatic Labeling

After receiving data, the system automatically labels classification labels in accordance with the level of sensitivity and regulatory relevance. Depending on machine learning models and preset policy rules, the data are separated into the range of categories, namely, public, internal, confidential, or highly sensitive. Automatic labeling also allows the organization to have a consistency in classification, and reduction of human error. These tags are used as the basis of implementing the right access control, encryption, and retention regulations. Labels are updated by means of continuous re-evaluation mechanisms, in case there are changes in the content and circumstantial context of the data.

3.4.3. Consent Verification

Consent verification makes sure that the data processing processes are within accurate user rights and legal guidelines. The system compares the records of given consents to the real data use in order to ensure processing purposes are not obsolete. It verifies things like consent scope and expiration status, withdrawal requests and regulatory restrictions. In case of inconsistencies, the alerts are created to be remediated. The presence of automated consent verification eliminates the threat of illegal processing and reinforces the adherence to privacy laws, like GDPR.

3.4.4. Risk Computation

The system then uses a privacy risk score, which depends on the sensitivity of data, trends in its access, how often accessed by a third party and regulatory requirements after the classification and consent validation. This step will combine a combination of risk indicators into a single assessment which mirrors the possibility of compliance exposure. The dynamic recalibration mechanisms provide risk scores to be adjusted to the processing activities or thriving risks. The urgency and nature of enforcement measure can be determined by the calculated risk level.

3.4.5. Policy Enforcement

Risk evaluation is translated into computerized control measures. The system can either block user access, activate encryption controls, implement data downsizing, or launch remediation processes depending on the risk level. The regulatory standards and internal governance policies are in congruence with enforcement rules. Automated enforcement also minimizes the time taken when responding and minimizes chances of human oversight failure.

3.4.6. Audit Logging

Audit captures records all compliance activities, such as data access events, classification decisions, risk assessment and enforcement activities. These are logs that make the internal reviews transparent and traceable to the external regulatory reviews. Detailed audit trails increase accountability, aid in investigation of incidents and indicate compliance with regulations. Trend analysis and compliance Boosting Long-term compliance monitoring Aids such as trend analysis are also facilitated by continuous logging.

4. Results and Discussion

4.1. Experimental Setup

The model of the experiment was created to assess the efficacy of the proposed compliance automation framework with reference to simulated enterprise datasets. [18,19] The datasets were designed to be similar to real organization worlds and contained both structured data like customer profiles, transaction history, employee data, and consent data and unstructured data like email, policy documents, and support tickets. The data were invariably destined to contain sensitive attributes, to test the strength of the classification and risk detection models. Different access patterns, inconsistent consent and possible violation of compliance were also included in the simulation in an effort to simulate a real-world risk situation in relation to regulation. The assessment was based on four main performance measures namely: accuracy, precision, recall, and rate of compliance improvement. The overall correctness of the data classification and anomaly detection models in identifying sensitive data and a

possible violation was measured through accuracy. Precision measured the percentage of compliance risks correctly identified within all the cases flagged that helps to determine the weakness of the system in the reduction of false positives.

The width of recall measured the fraction of actual compliance risks that became effectively identified which made sure that the most crucial violations get not neglected. Along with traditional machine learning metrics, the indicator of compliance improvement rate was also proposed as a domain-specific metric. This measure determined the percentage decline in compliance gaps against a rule-based system, which was obtained by applying automated enforcement mechanisms using a baseline system. All the experiments were done in controlled simulation conditions and several training and testing cycles were repeated to achieve consistency and reproducibility. It allowed the complete verification of the classification performance, the use of the anomaly detection tools, and the overall assessment of the regulatory risk reduction feature within a realistic enterprise environment.

4.2. Performance Results

Table 1. Performance Results

Metric	Rule-Based System (%)	Proposed ML System (%)
Data Classification Accuracy	78%	94%
Consent Verification Accuracy	72%	91%
Violation Detection Rate	65%	89%
False Positive Reduction	40%	18%
Automated Compliance Coverage	55%	93%
Risk Prediction Accuracy	70%	92%

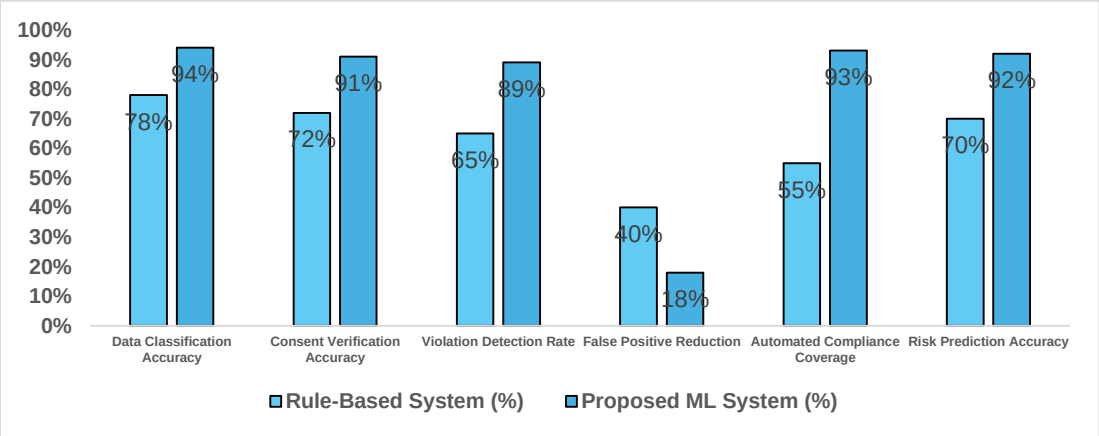

Figure 5. Performance Results

4.2.1. Data Classification Accuracy

This proposed machine learning system attained a data classification of 94 percent, and it is much higher than the rule-based system which had attained 78 percent. This advancement proves the efficiency of the integration of NLP strategies with deep neural networks to detect sensitive and controlled information. The ML-based model in contrast to rule-based approaches which are based on the assumption of pre-defined keywords or patterns can understand contextual meaning and semantic relationship in structured and non-structured data. It thus minimises the probability of misclassification and enhances the overall detection reliability with a wide variety of enterprise data.

4.2.2. Consent Verification Accuracy

The accuracy in consent verification increased in the proposed system of the ML by 91 percent compared to the rule-based system of 72 percent. The conventional system was plagued with compound consent cases, which include conditional authorizations, expired consent and partial withdrawals. The ML-powered structure includes a contextual analysis and dynamic rule mapping, which allows being more precise when trying to tie together the use of data and consent records stored by the company. This enhancement enhances compliance with the regulations people ensure that the processing operation is more consistent with the user permissions.

4.2.3. Violation Detection Rate

The rate of violation detection levels rose significantly to 89 percent as compared to 65 percent. The rule based system lacked the ability to detect complex or indirect violations of rules especially of behavioral anomaly or policy slips. By comparison, the suggested ML system incorporates the methods of anomaly detection that breaks down abnormal access patterns and dangerous processing behaviors. This will result in more extensive detection of possible violations and proactive compliance control.

4.2.4. False Positive Reduction

The false positive rates in the proposed system were drastically lower than the rule based model where the rate worsened to 40 percent. Traditional systems have very high false positives, which tends to overload manual review work and lead to unnecessary alerts. The ML-based methodology enhances accuracy through learning decision boundaries based on past trends and refining them to reduce false risk indications and streamline its operations.

4.2.5. Automated Compliance Coverage

There was an extreme improvement in automated compliance coverage to 93 percent. The system based on rules necessitated a large number of manual operations during the updating and enforcement of policies. Comparatively, the suggested system incorporates automated categorization, risk rating, and policy coordination, which allows the implementation of wider and more uniform enforcement of enterprise data assets. This increase in the coverage will guarantee that there is increased percentage of data processing activities that are automatically observed and managed.

4.2.6. Risk Prediction Accuracy

The accuracy of risk prediction in the rule-based system was 70 percent, whereas in the ML-based one, the accuracy had risen to 92 percent. The improved model uses a variety of risk instruments including data sensitivity, anomaly detecting results and consent validity signals. The suggested system is more trustworthy and more context-sensitive risk measurements obtained through adaptive weighting and dynamic recalibration. This enhancement facilitates enhanced prioritization of compliance interventions and risk strategic reduction.

4.3. Discussion

The experiment evidence is clear evidence that the suggested machine-based-learning-informed compliance framework presents considerable performance gains in comparison to the conventional rule-based systems. Most importantly, the system has resulted in an increase in accuracy in data classification by 16 percent. This improvement can be explained by the fact that hybrid NLP and deep learning technology allow understanding sensitive information depending on the context and not using predefined keywords or official rules, only. The better quality of classification directly enhances subsequent operations, including consent validation and risk scoring, which minimizes the chances of misclassification of data assets by minimizing the risk of miscontinuity of the overall governance. Besides, the framework showed that the violation detection capability increased by 24 percent. Achieving quality results on this in traditional rule-based solutions is often difficult because these systems may not capture complex, or emerging compliance risks, especially around behavioral anomaly or indirect policy violation. The proposed system will be able to detect the slight deviation in data access and processing actions by introduction of anomaly detection algorithms and probabilistic risk modeling.

This proactive detection feature helps in improving the performance of the organization to act proactively to possible regulatory breaches before they are turned into serious compliance breaches or monetary penalties. The greatest operationally influential outcome, perhaps, is the 38 percent boost in automated compliance coverage. Automation coverage shows the percentage of compliance operations that may be measured and executed automatically. The architecture suggested combines the process of data discovery, classification, calculation of risks, and policy orchestration into a single workflow, increasing the range of automated enforcement implemented in the enterprise systems. Higher automation does not only result in the minimization of administrative overhead but also proper policy application across jurisdictions and departments. Taken together, these advantages indicate that the ML-based solution increases accuracy, detection, and scalability of the operation, which makes it one of the strong solutions to the current regulatory landscape.

5. Conclusion

The provided research is a groundbreaking framework based on machine learning that allows achieving end-to-end regulation adherence to the requirements of GDPR and DMA in enterprise contexts. The proposed architecture has incorporated smart data discovery, automatic classification, anomaly detection, privacy risk scoring, compliance coordination, and federated learning into one coherent and scalable platform. The framework also increases greatly the accuracy, responsiveness and

coverage of compliance operations by substituting non-adaptive rule-based mechanisms with adaptive machine learning models. The hybrid data classification model is better than the others because it is effective to identify sensitive and regulated data, and suspicious anomaly detection systems enhance the system to identify irregular access patterns and possible policy violations. More so, the formal privacy risk scoring module offers measurable and dynamic risk evaluation that reinforces prioritized control procedures and sound governance choices. One of the main contributions of this study is the combination of federal learning that allows training models without centralizing sensitive data. The strategy can save data privacy among organizational units and jurisdictions at the same time share the benefits of cooperative model enhancement.

The feasibility and applicability of ML-based compliance orchestration are confirmed by the findings of the experiment, which were based on simulated enterprise data until the beginning. There was a substantial increase in the quality of classification, the rate of violation detection, automation coverage, and general risk prediction consistency. These results indicate that smart automation will be able to greatly minimise false positives, enhance efficiency in operations, and enhance regulatory congruence in intricate enterprise ecosystems. In spite of these developments, more improvements can be done. The next steps in research could be developing blockchain-based audit trails to increase the levels of transparency, immutability, and trust regarding compliance reporting. Blockchain can offer immutable logging systems, which will enhance responsibility in case of regulatory auditing. Also, it might be added that some form of reinforcement learning could be introduced to facilitate policy adaptation to changes. This would enable the system to dynamically adapt the enforcement strategies according to changing risk trends, changes in regulations and organizational changes. Through feedback loops and never-ending self-improvement based on the results of learning, the framework would be able to advance to self-optimizing compliance governance. All in all, this study forms a compelling basis of smart, scalable, and regulation-compliant compliance systems in the contemporary digital businesses.

References

[1] Johnson, M., Jain, R., Brennan-Tonetta, P., Swartz, E., Silver, D., Paolini, J., ... & Hill, C. (2021). Impact of big data and artificial intelligence on industry: developing a workforce roadmap for a data driven economy. Global Journal of Flexible Systems Management, 22(3), 197-217.

[2] Yablonsky, S. (2019). Multidimensional data-driven artificial intelligence innovation. Technology innovation management review, 9(12), 16-28.

[3] Breaux, T. D., Antón, A. I., & Doyle, J. (2008). Semantic parameterization: A process for modeling domain descriptions. ACM Transactions on Software Engineering and Methodology (TOSEM), 18(2), 1-27.

[4] Maxwell, J. C., & Antón, A. I. (2009, August). Developing production rule models to aid in acquiring requirements from legal texts. In 2009 17th IEEE International requirements engineering conference (pp. 101-110). IEEE.

[5] Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. IEEE Internet Computing, 21(6), 58-62.

[6] Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. Information systems research, 15(4), 336-355.

[7] Toubiana, V., Narayanan, A., Boneh, D., Nissenbaum, H., & Barocas, S. (2010, March). Adnostic: Privacy preserving targeted advertising. In Proceedings Network and Distributed System Symposium.

[8] Sun, Y., Liu, J., Wang, J., Cao, Y., & Kato, N. (2020). When machine learning meets privacy in 6G: A survey. IEEE Communications Surveys & Tutorials, 22(4), 2694-2724.

[9] Shokri, R., & Shmatikov, V. (2015, October). Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC conference on computer and communications security (pp. 1310-1321).

[10] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2, pp. 1-800). Cambridge: MIT press.

[11] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

[12] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

[13] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

[14] Voigt, P., & Von dem Bussche, A. (2017). The eu general data protection regulation (gdpr). A practical guide, 1st ed., Cham: Springer International Publishing, 10(3152676), 10-5555.

[15] Butin, D., & Le Métayer, D. (2015, May). A guide to end-to-end privacy accountability. In 2015 IEEE/ACM 1st International Workshop on TEchnical and LEgal aspects of data pRivacy and SEcurity (pp. 20-25). IEEE.

[16] Governatori, G. (2009, December). Rule-based versus principle-based regulatory compliance. In Legal Knowledge and Information Systems: JURIX 2009, the Twenty-second Annual Conference (Vol. 205, p. 37). IOS Press.

[17] Vagnoni, G., Eisenbarth, M., Andert, J., Sammito, G., Schaub, J., Reke, M., & Kiausch, M. (2019). Smart rule-based diesel engine control strategies by means of predictive driving information. International Journal of Engine Research, 20(10), 1047-1058.

[18] Hossain, E., Khan, I., Un-Noor, F., Sikander, S. S., & Sunny, M. S. H. (2019). Application of big data and machine learning in smart grid, and associated security concerns: A review. Ieee Access, 7, 13960-13988.

[19] Esayas, S., & Mahler, T. (2015). Modelling compliance risk: a structured approach. Artificial Intelligence and Law, 23(3), 271-300.

[20] Lavanya, P. M., & Sasikala, E. (2021, May). Deep learning techniques on text classification using Natural language processing (NLP) in social healthcare network: A comprehensive survey. In 2021 3rd international conference on signal processing and communication (ICPSC) (pp. 603-609). IEEE.
[21] Salur, M. U., & Aydin, I. (2020). A novel hybrid deep learning model for sentiment classification. IEEE Access, 8, 58080-58093.
[22] Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. Knowledge-Based Systems, 216, 106775.