Smart Governance among Smart Cities for Legal Consideration to International Data Migration in Cloud Using Machine Learning , Nlp and Blockchain Smart Contract

Komala R , Arun Kumar B.R. * , Mahadeshwara Prasad , Shreyas A

Posted Date: 15 August 2024

doi: 10.20944/preprints202408.1028.v1

Keywords: cloud computing; machine learning; natural language processing; blockchain; smart contract; GDPR; CCPA; international data transfer policies; data protection regulation

 doi:10.20944/preprints202408.1028.v1

Article

Smart Governance among Smart Cities for Legal Consideration to International Data Migration in Cloud Using Machine Learning , Nlp and Blockchain Smart Contract

Komala R [1,‡], Arun Kumar B.R. [2,*,†,‡], Mahadeshwara Prasad [3,‡] and Shreyas A [4,‡]

[1] Ph.D Research Scholar, Dept. of MCA, VTU RC, BMS Institute of Technology & Management, and Assistant Professor, Department of Computer Applications, M S Ramaiah Institute of Technology, Bengaluru, India;

[2] Professor, Department of Computer Science & Engineering & Research Supervisor, Department of MCA, BMS Institute of Technology & Management, Yelahanka, Bengaluru, India

[3] UG Scholar, Department of Computer Science & Engineering, BMS Institute of Technology & Management, Yelahanka, Bengaluru, India

[4] UG Scholar, Department of Computer Science & Engineering, Sai Vidya Institute of Technology, Rajanukunte, via Yelahanka, Bengaluru, 560064, India

* Correspondence: arunkumarbr@bmsit.in; Tel.: 91-9886008210 : Arun Kumar B.R.

† Current address: Department of Computer Science & Engineering, BMS Institute of Technology and Management, Yelahanka, Bengaluru, India.

‡ These authors contributed equally to this work.

Abstract: Legal consideration hold significant importance in the cloud migration process, encompassing contractual arrangements, data sovereignty concerns, and liability matters. Organizations need to make sure that their contracts with cloud service providers (CSPS) cover important aspects such as data ownership, usage rights, and indemnification clauses. In the ever-changing world of smart cities, the importance of ensuring secure, compliant, and efficient data migration across international borders has become more crucial than ever. This paper introduces a new framework that combines natural language processing (NLP), and blockchain smart contracts to tackle the intricate legal issues involved in moving data across borders in cloud settings. The framework starts by utilizing an NLP model to ensure compliance with data protection regulations, such as GDPR, CCPA, and DPDPA, which are specific to the destination jurisdiction of the data. After confirming the verification, the smart contract initiates the data transfer process, securely recording metadata such as file hash, timestamp, and transfer details on the blockchain, guaranteeing transparency and immutability. After the transfer, an international vendor at the destination verifies the data against the relevant legal requirements, guaranteeing compliance before storing it in the destination cloud. By adopting this approach, we can ensure the legal validity of cross-border data transfers, while also promoting trust and accountability among all parties involved in smart city ecosystems. The findings indicate that this framework has the potential to greatly reduce the risks associated with data sovereignty, liability, and contractual obligations when moving data to the cloud.

Keywords: cloud computing; machine learning; natural language processing; blockchain; smart contract; GDPR; CCPA; international data transfer policies; data protection regulation

1. Introduction

The rapid expansion of smart cities has presented significant obstacles in handling and safeguarding data, particularly during the process of transferring data across borders. With the increasing interconnectivity of cities, ensuring the security, compliance, and efficiency of data transfers across borders has become a pressing issue. The implementation of data sovereignty, legal compliance, and privacy regulations like gdpr, ccpa, and dpdpa further complicate the process of data transfer. This research is crucial because it tackles these challenges by suggesting a framework that guarantees legal compliance, promotes transparency, and builds trust among stakeholders during international data migrations in cloud environments.

 doi:10.20944/preprints202408.1028.v1

NLP is a vital component of this framework as it automates the verification of legal agreements and regulatory compliance. Due to the intricate and multifaceted nature of international data protection laws, manually verifying compliance can be a time-consuming and prone-to-errors process. Our nlp model is specifically designed to parse and analyze legal texts, identify relevant regulations, and ensure that data migration requests comply with the legal requirements of the destination jurisdiction. This automation not only expedites the process but also minimizes the risk of non-compliance, thereby protecting the data migration process from legal complications.

Blockchain technology, in conjunction with smart contracts, is a fundamental component of our proposed framework. Blockchain's inherent qualities—unchangeability, transparency, and decentralized verification—make it an excellent choice for recording and monitoring data transfer activities. By implementing smart contracts, we automate the execution of data transfers once compliance is verified, guaranteeing that all conditions are fulfilled before initiating the process. The smart contract also keeps track of important information, like file hashes, timestamps, and transaction details, on the blockchain, ensuring that the audit trail is secure and cannot be altered. This not only guarantees the accuracy and reliability of the data but also builds trust among stakeholders by providing tangible proof of compliance and accountability throughout the data migration process.

The combination of natural language processing, blockchain, and smart contracts provides a robust solution that capitalizes on the unique capabilities of each technology. Nlp improves the precision and speed of compliance checks, blockchain guarantees the safety and openness of data transactions, and smart contracts automate and enforce legal agreements. By integrating these technologies, our framework tackles the legal and technical obstacles associated with moving data across borders while establishing a more dependable and secure system. By adopting a comprehensive approach that encompasses legal compliance and data security, smart cities can mitigate the risks of legal violations and data breaches while ensuring a flexible and adaptable solution that can meet the ever-changing demands of urban environments.

2. Related Work

The landscape of cloud storage and information migration has been appreciably explored in recent years, with a strong emphasis on improving security, performance, and consumer privacy including [1–25]. A good sized frame of studies has targeted at the implementation of superior automation strategies, inclusive of cryptograpy and smart contract, to shield touchy facts and optimize garage approaches. Moreover, the adoption of blockchain era and clever contracts has won traction as a means to offer transparent, tamper-proof records control and get admission to control mechanisms. These innovations goal to cope with the developing worries over facts breaches, secure data transfer, and compliance with stringent regulatory frameworks. This section critiques key contributions and findings in the field, highlighting the combination of NLP model and blockchain technology, and positioning our work in the broader context of at ease and green cloud storage answers.

Table 1. Related reference findings in our investigation.

Author	Citation	Title	Objectives	Findings
Padmanaban (2024)	[3]	Revolutionizing Regulatory Reporting through AI/ML: Approaches for Enhanced Compliance and Efficiency.	A Machine learning approach to provide a novel solution for reporting mandates in financial institutions.	By leveraging ML technique a framework can be made to ensure security polices before uploading or sending data.

 doi:10.20944/preprints202408.1028.v1

Shabnam Hassani, Mehrdad Sabetzadeh, Daniel Amyot and Jain Liao (2024)	[4]	"Rethinking Legal Compliance Automation: Opportunities with Large Language Models.	Employing automation strategies for compliance analysis using Large Language Models	Usage of NLP model for automation to check International data policy compliance.
Prajakta Sudhir Samant (2024)	[5]	Leveraging AI for enhanced compliance with global data protection regulations in cloud computing environments.	To analyze the critical role of compliance and regulatory adherence in cloud computing, with a focus on global data protection regulations like GDPR, CCPA, and PIPEDA.	Ensuring global data protection regulation for international data transfer without any human interaction.
Sanjeev Prakash, Jesu Narkarunai Arasu Malaiyappan, Kumaran Thirunavukkarasu and Munivel Devan (2024)	[6]	Achieving Regulatory Compliance in Cloud Computing through MLE.	To investigate the role of machine learning in enhancing regulatory compliance within cloud environments by automating tasks, improving security, and increasing reporting accuracy.	Improving the efficiency and accurary of the NLP model for automation compliance check.
Cristòfol Daudén-Esmel, Jordi Castellà-Roca, Alexandre Viejo (2024)	[8]	Blockchain-based access control system for efficient and GDPR-compliant personal data management	To develop a lightweight blockchain-based platform for GDPR-compliant personal data management, enabling Service Providers to transparently demonstrate consent and comply with regulations.	Leveraging smart contracts for data movement initiation after successful compliance check.
Konstantinos Demertzis, Konstantinos Rantos, Lykourgos Magafas, Charalabos Skianis and Lazaros Iliadis (2023)	[9]	A Secure and Privacy-Preserving Blockchain-Based XAI-Justice System.	To propose a framework that integrates AI innovations, including NLP, ChatGPT, and blockchain, to enhance the efficiency, transparency, and impartiality of judicial determinations.	Integration of NLP and blockchain technology to improve the efficiency and security of the system.

 doi:10.20944/preprints202408.1028.v1

Richmond Y. Wong, Andrew Chong, R. Cooper AspegrenAuthors Info and Claims (2023)	[10]	Privacy Legislation as Business Risks: How GDPR and CCPA are Represented in Technology Companies' Investment Risk Disclosures.	To investigate how major technology companies translate GDPR and CCPA into business risks in documents created for investors, focusing on the implications of privacy legislation on their operations.	How our system can be useful in business fields for companies which relay on global data protection policies.
O. A. Cejas, M. I. Azeem, S. Abualhaija and L. C. Briand (2023)	[11]	NLP-Based Automated Compliance Checking of Data Processing Agreements Against GDPR.	To develop an automated solution using natural language processing (NLP) for checking the compliance of Data Processing Agreements (DPAs) with GDPR requirements, aiming to streamline the compliance verification process.	Leveraging NLP for compliance check of global data protection policies like GDPR, CCPA and DPDPA.
Tom, J., Adigwe , W., Anebo, N., and Bukola (2023)	[12]	Automated Model for Data Protection Regulation Compliance Monitoring and Enforcement.	To develop an automated compliance and enforcement model using Semantic Web technologies and Ontology for monitoring adherence to data protection regulations, specifically targeting the Nigerian Data Protection Regulation (NDPR).	Automation model for data protection regulation like NLP for multiple country's policies.
Filippo Lorè, Pierpaolo Basile, Annalisa Appice, Marco de Gemmis, Donato Malerba and Giovanni Semeraro (2023)	[13]	An AI framework to support decisions on GDPR compliance.	To design and implement the INTREPID AI-based framework for automating the GDPR compliance check of public documents within the Italian Public Administration, focusing on Italian language processing.	Automation system for a secure compliance check of GDPR policies for user's data.

 doi:10.20944/preprints202408.1028.v1

Haris Ahmad, Gagangeet Singh Aujla (2023)	[15]	GDPR compliance verification through a user-centric blockchain approach in multi-cloud environment.	To develop a user-centric, blockchain-based framework for managing and verifying GDPR compliance in cloud-hosted web applications, focusing on transparent and immutable logging of data operations via smart contracts.	Integration of smart contract with NLP model to enhance more security and privacy of user data for global data policies compliance.
L. Wang, Z. Guan, Z. Chen and M. Hu (2023)	[17]	Enabling Integrity and Compliance Auditing in Blockchain-Based GDPR-Compliant Data Management.	To develop a blockchain-based data management framework that ensures both semantic consistency and data integrity in compliance with GDPR, enabling transparent data operations and inspections.	Usage of blockchain to ensure data integrity and evidence of data transfer from one cloud to another.
Masoud Barati, Kwabena Adu-Duodu, Omer Rana, Gagangeet Singh Aujla and Rajiv Ranjan (2023)	[18]	Compliance Checking of Cloud Providers: Design and Implementation.	To develop and verify a formal model for data usage requests in cloud composite services, ensuring compliance with GDPR obligations such as user consent, data access, and data transfer.	Providing an environment for user consent, data access, data transfer which ensures users a secure and privacy preserving trust with in the system.
Dara Hallinan, Alexander Bernier, Anne Cambon-Thomsen, Francis P. Crawley, Diana Dimitrova, Claudia Bauzer Medeiros, Gustav Nilsonne, Simon Parker, Brian Pickering and Stéphanie Rennes (2021)	[23]	International transfers of personal data for health research following Schrems II: a problem in need of a solution,	To analyze the impact of the Schrems II decision on the transfer of personal data for health research between the EU and third countries, particularly in the context of the COVID-19 pandemic.	Impact of transferring crucial data from one country to another by agreement of all global data security policies.

 doi:10.20944/preprints202408.1028.v1

Mpyana Mwamba Merlec, Youn Kyu Lee, Seng-Phil Hong and Hoh Peter (2021)	[24]	A Smart Contract-Based Dynamic Consent Management System for Personal Data Usage under GDPR.	To develop and implement a smart-contract-based dynamic consent management system using blockchain technology, enabling individuals to control personal data collection and usage in compliance with GDPR.	Leveraging smart contract as it provide immutability, So that users can have control over their personal data.

The extensive literature analysis has lead to identify several research gaps. There is still a considerable research gap in developing a single framework that smoothly integrates all of these components, despite major breakthroughs in automating data migration operations, guaranteeing legal compliance with NLP, and incorporating blockchain for safe transfers. Studies conducted today frequently concentrate on particular facets, such as GDPR compliance, the use of smart contracts for particular laws, or the use of natural language processing (NLP) in data policy verification. But there is still more work to be done to develop a comprehensive solution that tackles the challenges of cross-border data transfers from all angles. These challenges include real-time compliance verification across multiple jurisdictions, adaptive NLP models that can change with the law, and blockchain's potential to improve data sovereignty and transparency. Subsequent investigations could concentrate on creating an integrated framework that guarantees cross-border data transfers' technical and legal stability as well as their flexibility and scalability in order to accommodate changing international data regulations.

3. Novelty of the work

The suggested framework signifies a notable progress in the domain of international data transfer, especially when applied to smart cities. The uniqueness of this project lies in the seamless combination of machine learning, natural language processing, blockchain technology, and smart contracts to tackle the intricate legal and technical obstacles encountered in cross-border data transfers.

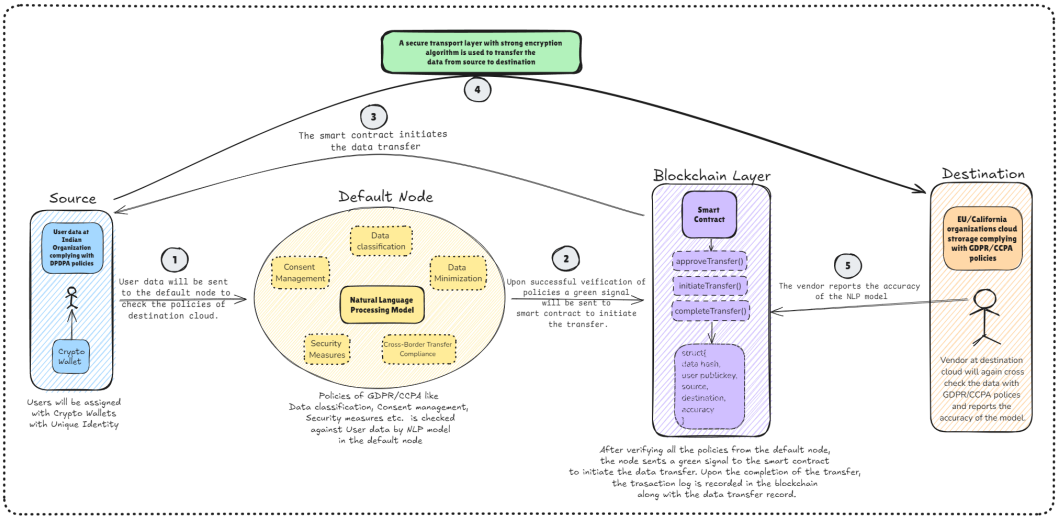

Figure 1. Novel architecture of secure international data transfer.

Firstly, the use of nlp to automatically verify compliance with diverse and often conflicting international data protection regulations is a groundbreaking approach. This model guarantees that

 doi:10.20944/preprints202408.1028.v1

data migrations adhere to legal requirements while also streamlining the process, reducing the time and human effort typically involved in verification, thereby minimizing the chances of human error and non-compliance with the law.

Additionally, the incorporation of blockchain technology to establish a transparent and unchangeable record of data transactions is a groundbreaking application that improves trust and accountability during the data migration process. By capturing important metadata on the blockchain, the framework establishes an unalterable audit trail, ensuring the highest level of security and traceability.

Lastly, the automation of data transfer processes through smart contracts is an innovative feature that guarantees that all legal conditions are fulfilled before the transfer is initiated. This not only simplifies the migration process but also ensures automatic compliance, minimizing the chances of legal conflicts and safeguarding the interests of all parties involved.

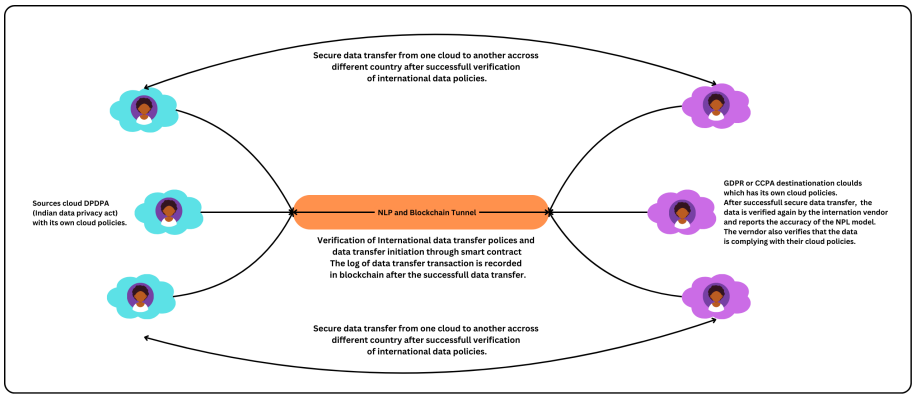

Figure 2. Novel architecture of NLP model and Smart contract tunnel.

4. Legal Consideration for International Cloud Data Transfer.

Inorder to provide a effective secure and privacy preserving result using of NLP model with blockchain technology is necessary. As there is a very less human interaction with the user data for verification of International data policies it ensures data privacy for the users. As the technology does a lot of job there will be no much waiting or processing time for the data transfer from one country/city to another. Below modeling shows the layers and steps we have used inorder to provide a secure international data transfer.

4.1. Natural Language Processing Modelling.

A complex Natural Language Processing (NLP) model is used in the context of automating data policy verification to guarantee adherence to laws like GDPR and DPDPA. The first step of the procedure involves using an embedding layer to transform input data into high-dimensional embeddings that capture semantic information about each token. Positional encodings are added to transformer-based models in order to maintain the token order. Subsequently, these embeddings are processed by multi-head attention methods, allowing the model to focus on multiple data points at once, thereby capturing intricate dependencies and relationships. A feedforward network is used to further modify the attention mechanism's output, increasing its representational capacity. Finally, the model is able to ascertain whether the data conforms with the given policies by using a softmax function to build a probability distribution across potential verification results. A signal is delivered to start the data transfer if the verification is successful, as shown by the model's output, guaranteeing that the procedure is carried out accurately and effectively without the need for human intervention.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 August 2024 doi:10.20944/preprints202408.1028.v1

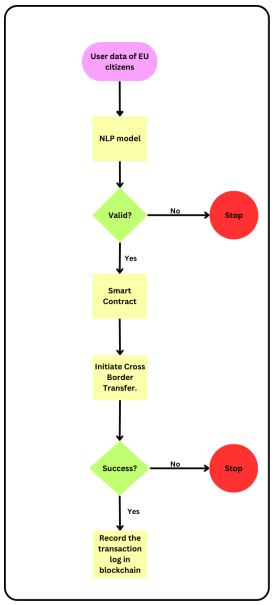

Figure 3. Flow control diagram of secure international data transfer.

4.1.1. Embedding Layer and Positional Encoding

The process begins with the input sequence $X = \{x_1, x_2, \ldots, x_n\}$, where each x_i represents a token or piece of data. The input sequence is first passed through an embedding layer to obtain a set of embedding vectors:

$$\mathbf{E} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\} = \text{Embed}(X)$$

where each $\mathbf{e}_i \in \mathbb{R}^d$ is a vector representation of the input token x_i.

For models such as transformers that require positional information, positional encodings are added to the embeddings:

$$\mathbf{z}_i = \mathbf{e}_i + \text{PE}(i)$$

where $\text{PE}(i)$ is the positional encoding vector for position i.

4.1.2. Multi-Head Attention

Next, the embeddings are processed through multi-head attention. The attention mechanism is defined as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}$$

where $\mathbf{Q} = \mathbf{W}^Q\mathbf{z}_i$, $\mathbf{K} = \mathbf{W}^K\mathbf{z}_i$, and $\mathbf{V} = \mathbf{W}^V\mathbf{z}_i$ are the query, key, and value matrices for the input $\mathbf{z}_i$, and d_k is the dimension of the key vectors.

4.1.3. Feedforward Layer

The outputs of the attention heads are concatenated and projected:

$$\mathbf{h}_i^{(1)} = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_h)\mathbf{W}^O$$

where $\mathbf{W}^O$ is the output projection matrix.

Following the attention mechanism, the output is passed through a feedforward network:

$$\mathbf{h}_i^{(2)} = \text{ReLU}(\mathbf{W}_1\mathbf{h}_i^{(1)} + \mathbf{b}_1)$$

$$\mathbf{h}_i^{(3)} = \mathbf{W}_2\mathbf{h}_i^{(2)} + \mathbf{b}_2$$

 doi:10.20944/preprints202408.1028.v1

where $\mathbf{W}_1$, $\mathbf{W}_2$, $\mathbf{b}_1$, and $\mathbf{b}_2$ are the weights and biases of the feedforward network.

4.1.4. Output Layer and Final Verification

Finally, the model produces an output probability distribution:

$$\mathbf{y} = \text{softmax}(\mathbf{W}_o \mathbf{h}_n^{(3)} + \mathbf{b}_o)$$

where $\mathbf{W}_o$ and $\mathbf{b}_o$ are the weights and biases of the output layer.

The final verification decision is given by:

$$\hat{y} = \arg\max(\mathbf{y})$$

If $\hat{y} = \text{Verified}$, the model sends a green signal to the smart contract to initiate the data transfer:

$$\mathbb{I}(\hat{y} = \text{Verified}) = 1$$

4.2. Smart Contract for Transfer Control.

Storing the encrypted metadata on the blockchain, the system guarantees the integrity and authenticity of the metadata, creating an unalterable record of the data. Furthermore, smart contracts are employed to regulate access control, guaranteeing that only authorized users can access and decrypt the files. By adopting this approach, organizations can not only enhance data security but also adhere to strict cloud policies that prioritize user privacy and data protection.

The modeling of the smart contract for initiating data transfer based on a green signal from the NLP model involves a sequence of actions and conditions ensuring the transfer complies with predefined policies.

4.2.1. Smart Contract State

Define the state of the smart contract as:

$$\text{State} = \{\text{dataController}, \text{approvedTransfers}\}$$

where 'dataController' is the authorized address, and 'approvedTransfers' is a mapping of data hashes to their approval status.

4.2.2. Green Signal from NLP Model

Let $\mathbb{I}(\hat{y} = \text{Verified}) = 1$ represent the green signal from the NLP model, indicating that the data complies with the necessary regulations and is ready for transfer.

4.2.3. Transfer Initiation

The transfer initiation is modeled as:

$$\text{emit TransferInitiated}(from, to, \text{dataHash})$$

This occurs when the smart contract receives a request to initiate a transfer. The function to initiate the transfer is:

$$\text{initiateTransfer}(\text{to}, \text{dataHash}) \rightarrow \text{emit TransferInitiated}(msg.sender, to, \text{dataHash})$$

4.2.4. Transfer Approval

Approval of the transfer by the data controller is modeled as:

$$\text{approvedTransfers[dataHash]} = \text{true}$$

 doi:10.20944/preprints202408.1028.v1

This update occurs when the 'approveTransfer' function is called, ensuring the transfer is authorized. The condition for approval is:

$$\text{approveTransfer}(\text{dataHash}) \quad \text{if} \mathbb{I}(\text{msg.sender} = \text{dataController}) = 1 \rightarrow \text{approvedTransfers[dataHash]} = \text{true}$$

and the contract emits:

$$\text{emit TransferApproved(dataHash)}$$

4.2.5. Transfer Completion

The transfer is completed only if the data hash has been approved:

$$\text{completeTransfer}(\text{from}, \text{to}, \text{dataHash}) \quad \text{if approvedTransfers[dataHash]} = \text{true}$$

Upon successful completion, the contract emits:

$$\text{emit TransferCompleted(from, to, dataHash)}$$

4.2.6. Combined Model

The overall model for the smart contract's operation with the NLP signal can be summarized as:

1. **Receive Green Signal:** $\mathbb{I}(\hat{y} = \text{Verified}) = 1$
2. **Initiate Transfer:** emit TransferInitiated$(msg.sender, to, \text{dataHash})$
3. **Approve Transfer:** approvedTransfers[dataHash] $=$ true if msg.sender $=$ dataController
4. **Complete Transfer:** emit TransferCompleted(from, to, dataHash) if approvedTransfers[dataHash] $=$ true

5. Experimental Results and Discussion

Performance analysis is essential when it comes to guaranteeing efficient data transfer and compliance verification using blockchain smart contracts and sophisticated NLP models. This analysis's main goal is to assess important metrics like the correctness of the NLP model, transfer latency, blockchain transaction throughput, and compliance verification time. Together, these metrics evaluate the effectiveness, dependability, and speed of the system as a whole, from transaction execution to model prediction. Through a thorough analysis of these variables, we can spot possible bottlenecks, boost efficiency, and make sure the system complies with legal standards all the while keeping a high level of operational effectiveness.

A robust computing setup is used to conduct this performance study in order to guarantee accurate and trustworthy results. The examination is performed on a machine that has an AMD Ryzen 7 4800H processor, 16GB of RAM, and Radeon graphics with a clock speed of 2.90 GHz. Operating on a 64-bit version of Windows 11, the system provides a stable environment in which to conduct performance testing and collect extensive data. The system's performance can be thoroughly assessed and benchmarked with this configuration, which offers the information required to improve the NLP model and smart contract implementations.

5.1. Training of NLP Model

The NLP model has been trained by giving inputs such as user's personal data like their identity information, credit card information etc. The model detects whether the data is confidential and the access control by authorized user. We have trained the model to check these two policies, further we will implement to verify more policies for GDPR, CCPA as well as India's DPDPA(Digital Personal Data Protection Act). The model will take users data as input and tells which policies are complying like Article 5(1)(f) - Integrity and Confidentiality, Article 32 - Security of Processing of GDPR and Section 1798.150 - Data Breach Liability, Section 1798.81.5 - Reasonable Security Procedures of CCPA.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 August 2024 doi:10.20944/preprints202408.1028.v1

TF-IDF vectorizer with a Naive Bayes are used in the model. The TF-IDF vectorizer transforms the raw text data into numerical form, where each word is represented by a score that reflects its importance in the document relative to the entire corpus. The Naive Bayes classifier then uses these numerical vectors to learn and predict labels for the text data.
Figure 4 shows a sample test dataset we used to train the model.
Note: Only some part of dataset is shown for example.

```
data = {
    "text": [
        "This is a normal text string.",
        encryption("This is an encrypted string."),
        "Credit card number: 1234 5678 9101 1121.",
        encryption("Credit card number: 1234 5678 9101 1121."),
        "User's password is stored securely.",
        encryption("User's password is stored securely.")
    ],
    "label": [
        "Non-compliant with GDPR Article 5(1)(f): Integrity and Confidentiality",
        "Compliant with GDPR Article 32: Security of Processing",
        "Non-compliant with GDPR Article 5(1)(f): Integrity and Confidentiality",
        "Compliant with GDPR Article 32: Security of Processing",
        "Non-compliant with GDPR Article 5(1)(f): Integrity and Confidentiality",
        "Compliant with GDPR Article 32: Security of Processing"
    ]
}
```

Figure 4. Sample dataset example used for training.

The procedure starts when a user enters their personal information, which is then safely sent via a REST API to an NLP model. The purpose of this model is to evaluate the data and guarantee adherence to pertinent data protection laws. The model particularly verifies GDPR compliance if the user plans to move their data to European nations. In contrast, the model evaluates data against CCPA regulations in the event that it is transported to California. The model then assesses whether the data management procedures adhere to the strict guidelines set forth in the relevant rules. Following compliance verification, the model gives the go-ahead for a smart contract to authorise and start the data migration process to the chosen area, guaranteeing that all legal requirements are are satisfied before the transfer occurs.

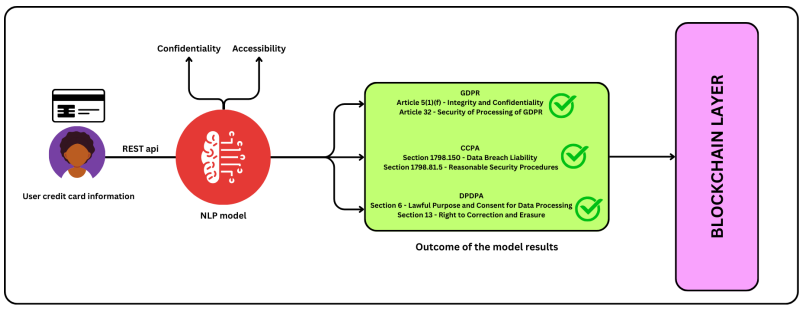

Figure 5. NPL model compliance check procedure.

5.2. NLP Model Accuracy

By comparing the model's output with the expected output for a specific set of test data, the accuracy of the NLP model is determined. The accuracy score gives information about the model's ability to accurately detect and confirm adherence to data policies.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \times 100$$

 doi:10.20944/preprints202408.1028.v1

Table 2. NLP Model Accuracy.

Test Input	Accuracy (%)	Expected Output	Actual Output
"Test Data 1"	95.0	True	True
"Test Data 2"	94.5	False	False
"Test Data 3"	96.0	True	True
"Test Data 4"	94.0	True	True
"Test Data 5"	95.5	False	False

Figure 6 shows the graphical representation of the accuracy of our model for various data.

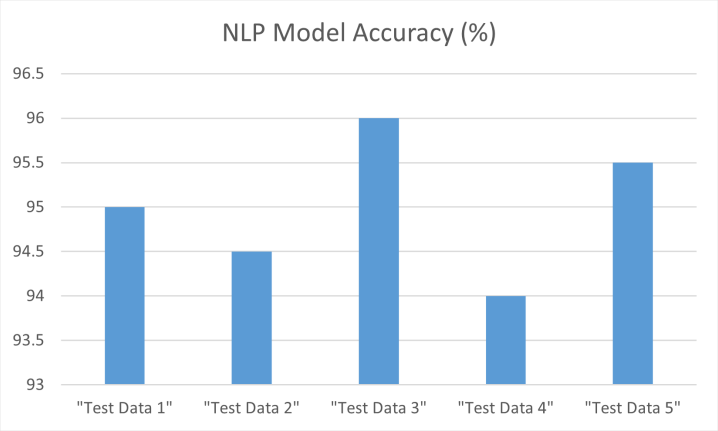

Figure 6. Graphical representation of NLP model accuracy.

5.3. Transfer Latency

The time it takes to start and complete a data transmission on a blockchain is known as transfer latency. This measure is essential for assessing how quickly and responsively the smart contract responds to requests for data transfers.

Table 3. Transfer Latency.

Data Input	Latency (ms)	Recipient Address
"Data Hash 1"	150	0xAddr1
"Data Hash 2"	160	0xAddr2
"Data Hash 3"	145	0xAddr3
"Data Hash 4"	155	0xAddr4
"Data Hash 5"	140	0xAddr5

Figure 7 represents the graphical representation of the latency of data transfer from source to destination.

 doi:10.20944/preprints202408.1028.v1

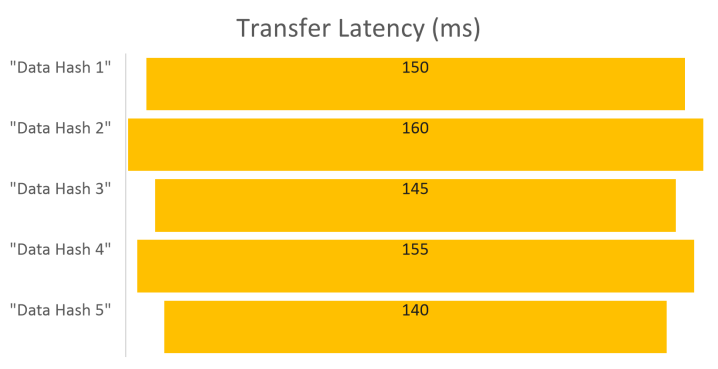

Figure 7. Graphical representation of latency in sec for the data to reach from source to destination.

5.4. Blockchain Transaction Throughput

The quantity of transactions that a blockchain network can handle in a specific amount of time is measured by blockchain transaction throughput. Increased throughput is a sign of a more effective network that can manage higher transaction volumes.

$$\text{Throughput} = \frac{\text{Number of Transactions}}{\text{Total Time (s)}}$$

Table 4. Blockchain Transaction Throughput.

Transaction Batch Size	Throughput (tx/s)	Total Transactions	Total Time (s)
Batch 1: 5 txs	10.5	5	0.48
Batch 2: 10 txs	11.0	10	0.91
Batch 3: 15 txs	10.2	15	1.47
Batch 4: 20 txs	10.8	20	1.85
Batch 5: 25 txs	11.2	25	2.23

Figure 8 represents the graphical representation of the scalability of the blockchain we have used. We have tested this by deploying the smart contract on polygon blockchain as it provides better scalability and less transaction fee which addresses many issues like latency and expense for out system.

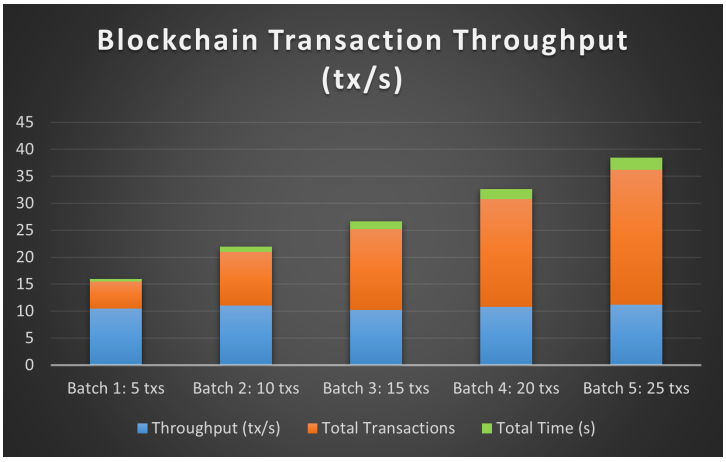

Figure 8. Graphical representation of Throughput.

 doi:10.20944/preprints202408.1028.v1

5.5. Compliance Verification Time

The amount of time the NLP model needs to determine if the incoming data complies with the pertinent data policies is known as the compliance verification time. This statistic aids in evaluating how well the model processes and reacts to compliance checks.

Table 5. Compliance Verification Time.

Compliance Data Input	Verification Time (ms)	Verification Result
"Policy 1"	200	Compliant
"Policy 2"	210	Non-compliant
"Policy 3"	190	Compliant
"Policy 4"	205	Compliant
"Policy 5"	195	Non-compliant

Figure 9 represents the graphical representation of time take for the verification process for checking compliance of the global data protection policies.

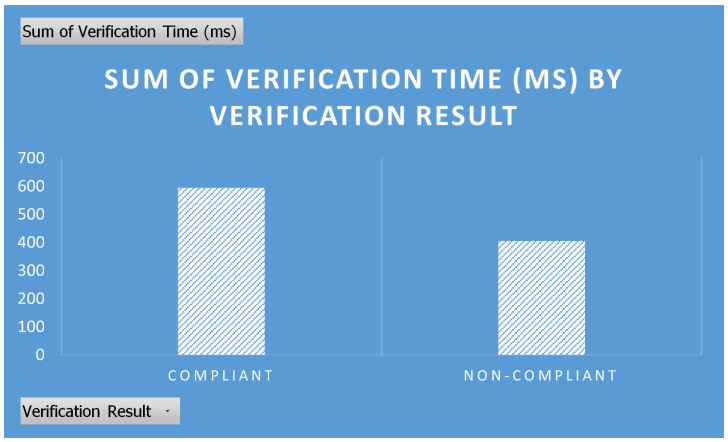

Figure 9. Graphical representation of Compliance Verification Time.

Important information about the efficacy and efficiency of the data transfer and compliance verification system is revealed by the performance study carried out on the NLP model and blockchain smart contract. We have developed a thorough grasp of the system's capabilities and limitations by analyzing parameters like compliance verification time, blockchain transaction throughput, transfer latency, and NLP model correctness. The investigation shows that the system operates well with precise compliance verification and effective data handling within the designated hardware and software environment. These results show that system performance can be further optimized, guaranteeing stable and dependable operations in practical applications.

6. Conclusions

This study uses the combined strengths of machine learning, natural language processing (NLP), blockchain technology, and smart contracts to propose a comprehensive framework that handles the complex issues of international data migration in the context of smart cities. The suggested method automates regulatory inspections, secures data transfers, and creates an unchangeable audit trail to guarantee legal compliance, improve transparency, and build confidence among stakeholders. Through the integration of these cutting-edge technologies, the framework not only reduces the hazards involved in cross-border data transfers, but it also develops a flexible and scalable solution to meet the ever-changing requirements of contemporary smart cities.

The findings show that the framework offers a strong solution that takes into account both technological and legal issues, greatly enhancing the effectiveness, security, and legal integrity of

 doi:10.20944/preprints202408.1028.v1

data migrations. The use of natural language processing (NLP) in automated policy verification, in conjunction with the immutability and openness of blockchain technology and the automation potential of smart contracts, represents a significant breakthrough in the field of global data governance.

Although the framework provides a strong answer, there are a few areas where additional improvements could bolster its potential. The NLP model might be improved by being expanded to include more languages and legal systems, which would enable wider application across many international locations. Furthermore, by using cutting-edge AI methods like deep learning, the NLP model's accuracy and versatility may be enhanced, making it capable of handling legal documents that are more intricate and nuanced.

Investigating interoperability between various blockchain platforms is another direction for future research. Interacting with numerous blockchain networks could improve the framework's scalability and flexibility as smart cities continue to develop. Incorporating privacy-preserving methods like homomorphic encryption or zero-knowledge proofs could also strengthen the security and confidentiality of data transfers, addressing issues with sensitive data.

Lastly, the framework might be expanded to incorporate machine learning and predictive analytics models that foresee and reduce possible security and legal issues before they materialize. In the event of international data migrations, the framework would provide even more protection and dependability by proactively addressing potential concerns. To sum up, this study establishes a solid foundation for upcoming advancements and breakthroughs in the field and paves the way for a fresh and practical approach to smart governance in smart cities.

References

1. Radhakrishnan Venkatakrishnan, Emrah Tanyildizi, and M. Abdullah Canbaz, "Semantic interlinking of Immigration Data using LLMs for Knowledge Graph Construction,". In *Companion Proceedings of the ACM on Web Conference 2024 (WWW '24)*, **2024**, doi: https://doi.org/10.1145/3589335.3651557 .
2. Jesu Narkarunai Arasu Malaiyappan, Sanjeev Prakash, Samir Vinayak Bayani and Munivel Devan, "Enhancing Cloud Compliance: A Machine Learning Approach,". *Advanced International Journal of Multidisciplinary Research*, **2024**, vol. 2, no. 2, doi: https://doi.org/10.62127/aijmr.2024.v02i02.1036.
3. Padmanaban, "Revolutionizing Regulatory Reporting through AI/ML: Approaches for Enhanced Compliance and Efficiency,". In *Journal of Artificial Intelligence General Science (JAIGS)*, **2024**, vol. 2, No. 1, pp. 71–90, doi: https://doi.org/10.60087/jaigs.v2i1.98.
4. Shabnam Hassani, Mehrdad Sabetzadeh, Daniel Amyot and Jain Liao, "Rethinking Legal Compliance Automation: Opportunities with Large Language Models,". In *arxiv*, **2024**, doi: https://doi.org/10.48550/arXiv.2404.14356.
5. Prajakta Sudhir Samant, "LEVERAGING AI FOR ENHANCED COMPLIANCE WITH GLOBAL DATA PROTECTION REGULATIONS IN CLOUD COMPUTING ENVIRONMENTS ". In *International Research Journal of Modernization in Engineering Technology and Science*, **2024**, vol.6, no. 4, pdf-link: https://www.irjmets.com/uploadedfiles/paper//issue_4_april_2024/53514/final/fin_irjmets1715711864.pdf.
6. Sanjeev Prakash, Jesu Narkarunai Arasu Malaiyappan, Kumaran Thirunavukkarasu and Munivel Devan, "Achieving Regulatory Compliance in Cloud Computing through ML". In *Advanced International Journal of Multidisciplinary Research*, **2024**,vol.2, no. 2, doi: https://doi.org/10.62127/aijmr.2024.v02i02.1038
7. Lillian Tsang "Transferring personal data out of the UK: The IDTA and UK addendum explained". *Artical*, **2024**, link: https://academic.oup.com/book/39321/chapter-abstract/350584629?redirectedFrom=fulltext.
8. Cristòfol Daudén-Esmel, Jordi Castellà-Roca, Alexandre Viejo, "Blockchain-based access control system for efficient and GDPR-compliant personal data management,". In *Computer Communications*, **2024**, vol. 214, pp. 67-87, DOI: https://doi.org/10.1016/j.comcom.2023.11.017.
9. Konstantinos Demertzis, Konstantinos Rantos, Lykourgos Magafas, Charalabos Skianis and Lazaros Iliadis, "A Secure and Privacy-Preserving Blockchain-Based XAI-Justice System". In *MDPI Information*, **2023**, vol.14, no. 9, doi: https://doi.org/10.3390/info14090477.

10. Richmond Y. Wong, Andrew Chong, R. Cooper AspegrenAuthors Info and Claims, "Privacy Legislation as Business Risks: How GDPR and CCPA are Represented in Technology Companies' Investment Risk Disclosures,". In *Accosiation for computing machinery*, **2023**, vol. 82, doi: https://doi.org/10.1145/3579515.
11. O. A. Cejas, M. I. Azeem, S. Abualhaija and L. C. Briand, "NLP-Based Automated Compliance Checking of Data Processing Agreements Against GDPR,". *IEEE Transactions on Software Engineering*, **2023**, vol. 49, no. 9, pp. 4282-4303, doi: https://doi.org/10.1109/TSE.2023.3288901.
12. Tom, J., Adigwe , W., Anebo, N., and Bukola, "Automated Model for Data Protection Regulation Compliance Monitoring and Enforcement,". In *International Journal of Computing, Intelligence and Security Research*, **2023**, vol. 2, no. 1, link: http://ijcsir.fmsisndajournal.org.ng/index.php/new-ijcsir/article/view/25.
13. Filippo Lorè, Pierpaolo Basile, Annalisa Appice, Marco de Gemmis, Donato Malerba and Giovanni Semeraro , "An AI framework to support decisions on GDPR compliance,". In *Springer Link*, **2023**, vol. 61, pp 541–568, doi: https://doi.org/10.1007/s10844-023-00782-4.
14. Song, J., Fu, H., Jiao, T. et al, "AI-enabled legacy data integration with privacy protection: A case study on regional cloud arbitration court,". In *Springer link, J Cloud Comp*, **2023**, vol. 12, no. 145, doi: https://doi.org/10.1186/s13677-023-00500-z.
15. Haris Ahmad, Gagangeet Singh Aujla, "GDPR compliance verification through a user-centric blockchain approach in multi-cloud environment,". In *Computers and Electrical Engineering*, **2023**, vol. 109, doi: https://doi.org/10.1016/j.compeleceng.2023.108747.
16. Bayani, S. V, Tillu, R and Jeyaraman, J, "Streamlining Compliance: Orchestrating Automated Checks for Cloud-based AI/ML Workflows,". In *Journal of Knowledge Learning and Science Technology*, **2023**, vol. 2, no. 3, doi: https://doi.org/10.60087/jklst.vol2.n3.p435.
17. L. Wang, Z. Guan, Z. Chen and M. Hu, "Enabling Integrity and Compliance Auditing in Blockchain-Based GDPR-Compliant Data Management,". In *IEEE Internet of Things Journal*, **2023**, vol. 10, no. 23, pp. 20955-20968 DOI: https://doi.org/10.1109/JIOT.2023.3285211.
18. Masoud Barati, Kwabena Adu-Duodu, Omer Rana, Gagangeet Singh Aujla and Rajiv Ranjan, "Compliance Checking of Cloud Providers: Design and Implementation,". In *Distributed Ledger Technologies: Research and Practice*, **2023**, vol. 2, no. 13, pp. 1-10. DOI: https://doi.org/10.1145/3585538.
19. Yunusa Simpa Abdulsalam and Mustapha Hedabou, "Security and Privacy in Cloud Computing: Technical Review,". In *MDPI, future internet*, **2022**, vol. 14, no. 1, doi: https://doi.org/10.3390/fi14010011.
20. Xu Ziyi, "International Law Protection of Cross-Border Transmission of Personal Information Based on Cloud Computing and Big Data". In *Wiley, Mobile Information System*, **2022**, doi: https://doi.org/10.1155/2022/9672693.
21. Yilun Zhou, Jianjun She, Yixuan Huang, Lingzhi Li, Lei Zhang andJiashu Zhang, "A Design for Safety (DFS) Semantic Framework Development Based on Natural Language Processing (NLP) for Automated Compliance Checking Using BIM: The Case of China". In *MDPI buildings*, **2022**,vol. 12, no. 6, doi: https://doi.org/10.3390/buildings12060780.
22. A. -J. Aberkane, G. Poels and S. V. Broucke, "Exploring Automated GDPR-Compliance in Requirements Engineering: A Systematic Mapping Study,". In *IEEE Access*, **2021**, vol. 9, pp. 66542-66559, link: https://doi.org/10.1109/ACCESS.2021.3076921.
23. Dara Hallinan, Alexander Bernier, Anne Cambon-Thomsen, Francis P. Crawley, Diana Dimitrova, Claudia Bauzer Medeiros, Gustav Nilsonne, Simon Parker, Brian Pickering and Stéphanie Rennes, "International transfers of personal data for health research following Schrems II: A problem in need of a solution,". In *European Journal of Humar Genetics*, **2021**, pp 1502–1509. DOI: http://dx.doi.org/10.1038/s41431-021-00893-y.
24. Mpyana Mwamba Merlec, Youn Kyu Lee, Seng-Phil Hong and Hoh Peter, "A Smart Contract-Based Dynamic Consent Management System for Personal Data Usage under GDPR,". In *MDPI sensors*, **2021**, vol. 21, no. 23, DOI: https://doi.org/10.3390/s21237994.
25. K. P. Joshi, L. Elluri and A. Nagar, "An Integrated Knowledge Graph to Automate Cloud Data Compliance,". In *IEEE Access*, **2020**,vol. 8, pp. 148541-148555, doi: https://doi.org/10.1109/ACCESS.2020.3008964.