ISSN: 2754-4516

Journal of Medicine and Healthcare

SCIENTIFIC Research and Community

Review Article

Open Access

Enterprise Data Security in Health Care Industry Using Data Science and Generative Adversarial Network to Implement Data Masking Techniques

Chandra Sekhar Veluru

USA

ABSTRACT

This white paper describes a comprehensive solution to enterprise data security in the healthcare industry. It achieves Enterprise Data Security in Healthcare using advanced Data Masking with Data Science and Generative Adversarial Network (GAN). Because of the growing digitalization of records in healthcare, it inevitably becomes very paramount to safeguard sensitive information belonging to the patients. In this paper, the authors have thrown light on the challenges in healthcare security, reviewed current data-masking techniques, and introduced one much different: Generative Adversarial Network (GAN). It improves data privacy, ensures conformance to regulations, and preserves data integrity and usability on healthcare data.

***Corresponding author**
Chandra Sekhar Veluru, USA.

Received: November 03, 2022; **Accepted:** November 10, 2022; **Published:** November 24, 2022

Introduction

The world is currently in a whirlwind transformation towards modernity, and healthcare has not been an exception with the adoption of electronic health records and digital healthcare systems. Many efficiency benefits and patient care gains are attained from these modern ways, but significant challenges, primarily data security, are thrown up. Protecting sensitive patient information has grown critical given rising cyber threats; therefore, healthcare data is most vulnerable with its ever-increasing value on the dark market.

The white paper on enterprise data security in the healthcare sector explores the use of data science and generative artificial intelligence to enable advanced data masking techniques. It thus proposes using sophisticated technologies, in particular GAN, to help support complete roofing for methods and solutions to ensure consistency of data privacy, sustain regulatory compliance in healthcare, and assure requisite health data integrity and usability.

The paper describes the present status of healthcare data security, the severe challenges for protecting organizations handling patient information, the traditional techniques of data masking, and an innovative approach using Generative Adversarial Network (GAN). It envisions a robust framework for healthcare organizations to improve data handling security, risks associated with breaches, and patients' trust and confidentiality within the digital healthcare ecosystem.

Importance of Data Security in Healthcare

The healthcare sector contains a vast pool of information, including personal and individual patients' data, medical history, treatment, and financial documentation. The values of this dataset are not only very significant when it comes to delivering quality healthcare services but are also jackpots for cybercriminals, who would be attracted to the exploitation of these data for financial or other ill intentions. This makes the healthcare sector a prime target for data breaches, exposing patients and an organization's reputation and economic stability to significant risks. The Cost of a Data Breach Report (IBM, 2020) underlined the seriousness of challenges in data security faced by the health sector. Healthcare organizations have the highest average cost per breached record compared to other industries. This highly frightening statistic, therefore, shows that effective data security measures were long overdue for protecting sensitive healthcare data [1-4].

Health data security is crucial in maintaining confidentiality, integrity, and availability. First, the rationale for the healthcare provider-patient relationship is maintaining patients' trust and confidence. By and large, patients share very intimate and personal information at a health institution, expecting that it should be treated discreetly by the health sector. Any breach of this trust due to a data security incident has drastic consequences that might undermine patients' confidence and drastically affect the credibility of the healthcare provider.

Regulatory compliance is the most important of all aspects of securing health information. Laws, such as HIPAA, have stringent requirements for ensuring adequate protection of patient data while placing punitive measures against an organization should a breach occur. Therefore, healthcare organizations should seek to ensure that they operate within the law to avoid legal suits, financial penalties, and other indispensable reputational damages. Data protection is of vital importance within any health environment. Healthcare organizations can significantly invest in cyber threat controls at the same time as maintaining more enhanced trust from patients with strict data security measures, adhering to a raft of regulations in force, and ensuring integrity and confidentiality in healthcare data by applying the most recent technologies developed for Data Science and GAN for Advanced Data Masking Techniques.

Current Challenges in Healthcare Data Security

Complexity of Healthcare Data

Healthcare data is sophisticated and incorporates various data types, such as structured data, unstructured data like medical images and notes, and semi-structured data like HL7 messages. Therefore, the diversity of health data makes it exceptionally awkward for any standard security practice to be applied uniformly to all data types. Moreover, health data often must be shared among departments, institutions, or even third-party service providers; data exposure risk under such circumstances becomes manifold.

Regulatory Compliance

The heart of the tightest regulatory requirements on patient information protection is where healthcare organizations must be. This is HIPAA in the United States, the General Data Protection Regulation in Europe, and the national data protection laws that call for rigid controls on handling and sharing healthcare data elsewhere. F fines and legal ramifications may be expensive, mainly when failure to comply is possible. Thus, compliance with these regulations without compromising operational efficiency becomes an enormous task for healthcare providers [5-8].

Insider Threats

Insider threats remain among the most significant healthcare data security risks, whether malicious or accidental. Any employee, contractor, or insider with legitimate access who misuses their privileges gladly goes ahead and causes a breach. According to the 2021 Verizon Data Breach Investigations Report, healthcare remains the industry with the most incidents related to insiders.

Cyber Threats

Ransomware, phishing, and advanced persistent threats are the most prevalent examples of cyber-attacks launched against healthcare organizations at ever-faster rates. The COVID-19 pandemic accelerated these risks by opening the vulnerabilities that companies inadvertently left open by moving overnight to remote work and telehealth services. A sophisticated threat detection capability and incident response should be implemented along with a continuous monitoring program layered at the core of protection from evolving threats.

Data Science for Data Masking in Healthcare

Data Science in Healthcare

Data science can change health care by applying statistical methods, algorithms, and machine learning techniques to extract useful information from complex data. Applications of data science range from predicting outbreak diseases using data analytics to personalized medicines and operational optimization in health care. As a whole, data science helps healthcare organizations deduce actionable insights concerning patient outcomes from data, improve their clinical decision-making processes, and thus enhance the overall quality of healthcare delivery.

Data Masking as a Security Measure

Data masking is a critical security practice that protects access to sensitive data. The process of masking accurate data ensures that if unauthorized people get control through a breach of defenses guarding the sensitive data, they will not be capable of getting sensitive data from the views or inputs presented to the end user. Data masking gives excellent value in non-production areas, specifically in testing, development, and training, where there is no need for actual data and presents an unnecessary security risk. This could be supplemented by traditional data masking methods, such as substitution, shuffling, and encryption, which are used daily but need help with scalability, performance, and data utility preservation.

Benefits of Artificial Intelligence for Data Masking

Artificial intelligence brings changes to the paradigm associated with data masking, touting several benefits that can potentially correct the inadequacies of the traditional methods of old in the following ways:

1. **Enhanced Data Privacy:** GAN ensures that masked data resembles accurate data greatly, thereby making it quite impossible to re-identify a person and the chances of breach of privacy.
2. **Scalability:** GAN models handle large, complex datasets with immunity, making them enterprise-appropriate for data masking.
3. **Data Utility:** The synthetic data generated by GAN captures the statistical and contextual characteristics of the source data, thereby embalming its usability for testing, development, and analytics purposes.
4. **Regulatory Compliance:** AI's effective obfuscation of sensitive information will enable healthcare organizations to comply with stringent data protection regulations regarding the integrity and confidentiality of health data. Data science and GAN will help further strong security, compliance with regulations, and information protection of patients while ensuring health data utility in its integrity for operation-critical and analytics purposes within healthcare institutions.

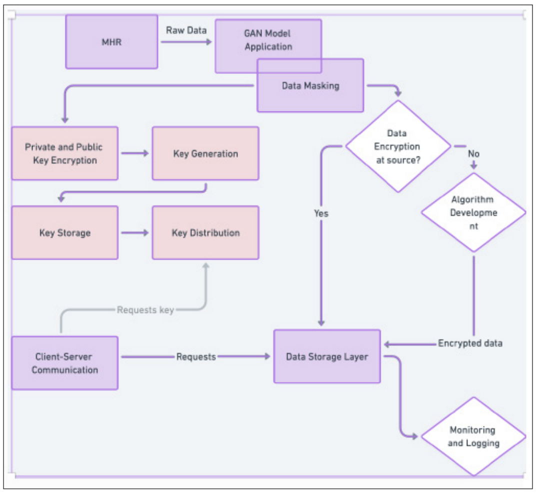
Synthetic Data Generation Design

The below modules are considered for implementation

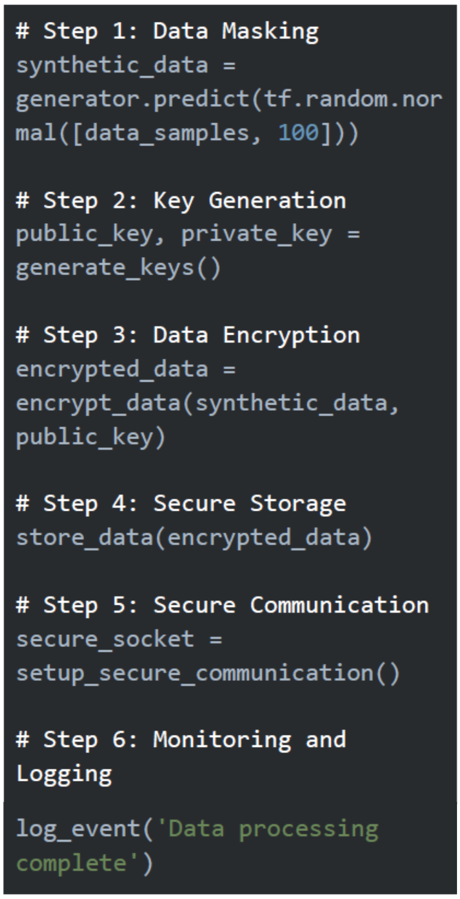
Module	Explanation
Data Masking Module	Utilize GAN and AI algorithms to generate synthetic data for MHR, including text and images. Implement masking techniques to anonymize sensitive information while preserving data utility.
Private and Public Key Encrypti on	Generate private and public keys for secure encryption and decryption processes. Implement client-server authentication using these keys to ensure safe data transmission.
Data Encrypti on at Source	Encrypt the data at the source using the public key before storing it in the data storage layer. Only authorized parties with the corresponding private key can decrypt and access the data.
Algorith m Development	Develop custom algorithms or masking techniques using GENAI and AI to enhance data security and privacy. Utilize advanced AI models to create innovative data masking strategies tailored for MHR.
Data Storage Layer	Store the encrypted synthetic MHR data securely in a data storage layer. Implement robust security measures to protect the stored data from unauthorized access.

Client-S erver Commu nication	Establish secure communication channels between clients and servers using private and public key encryption. Authenticate clients using their private keys and verify their identity with the corresponding public keys.
Monitori ng and Logging	Implement logging mechanisms to track data access, encryption, and decryption activities. Monitor system performance and security metrics to ensure data integrity and confidentiality.

Below is the high-level flowchart of the MHR record masking and storage:



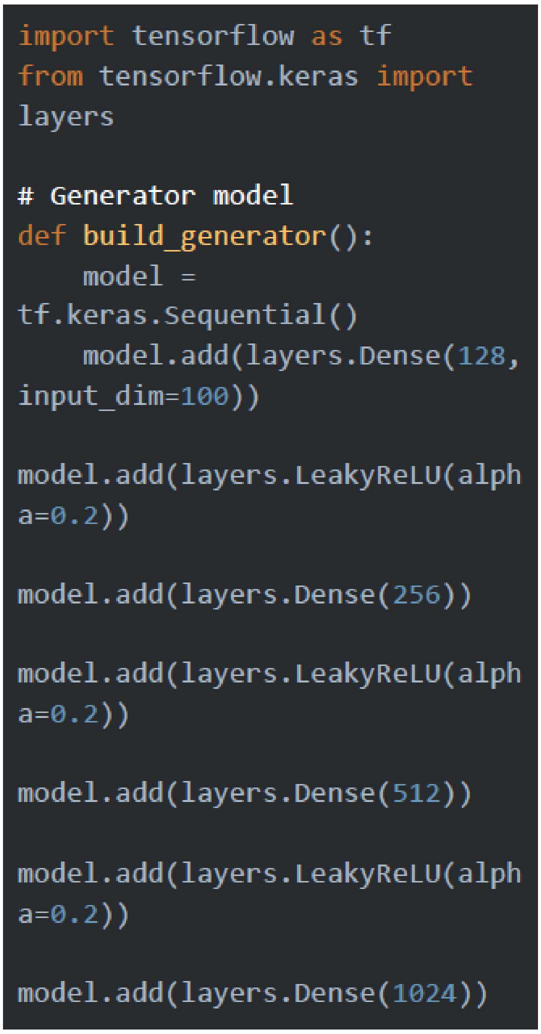
Trigger Script



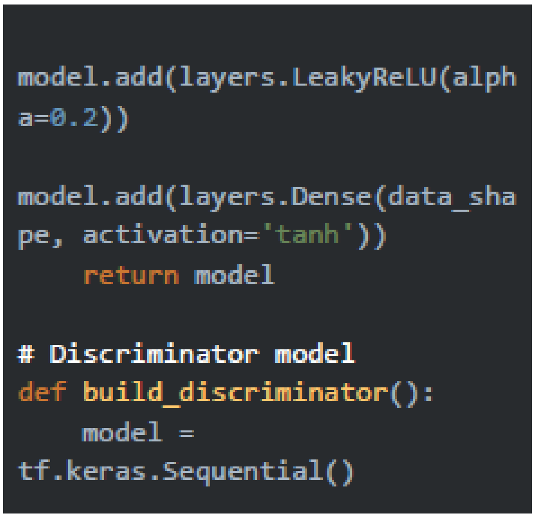
```
# Step 1: Data Masking
synthetic_data =
generator.predict(tf.random.nor
mal([data_samples, 100]))

# Step 2: Key Generation
public_key, private_key =
generate_keys()

# Step 3: Data Encryption
encrypted_data =
encrypt_data(synthetic_data,
public_key)

# Step 4: Secure Storage
store_data(encrypted_data)

# Step 5: Secure Communication
secure_socket =
setup_secure_communication()

# Step 6: Monitoring and
Logging

log_event('Data processing
complete')
```

GAN-based Data Masking Algorithm

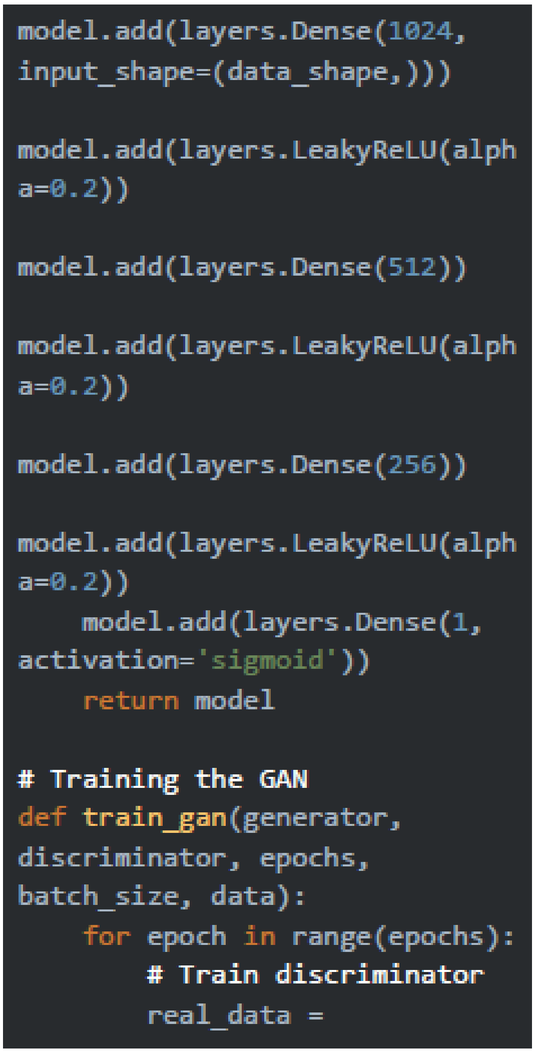
The script uses a Generative Adversarial Network (GAN) to create synthetic data that mimics the original MHR data. The generator model creates synthetic samples, and the discriminator model evaluates their authenticity, iteratively improving both models through training



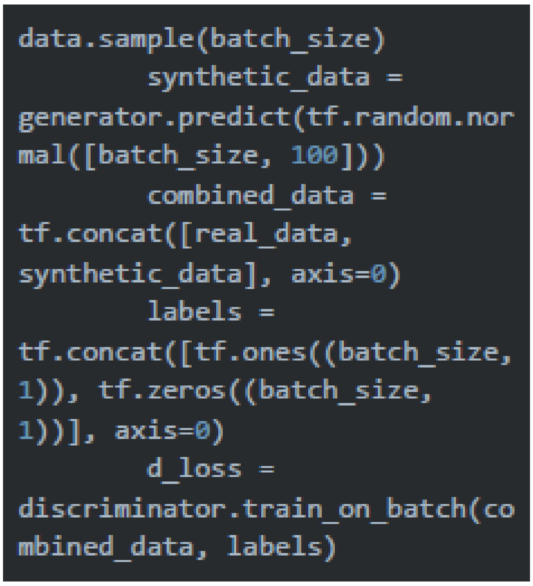
```
import tensorflow as tf
from tensorflow.keras import
layers

# Generator model
def build_generator():
    model =
tf.keras.Sequential()
    model.add(layers.Dense(128,
input_dim=100))

model.add(layers.LeakyReLU(alph
a=0.2))

model.add(layers.Dense(256))

model.add(layers.LeakyReLU(alph
a=0.2))

model.add(layers.Dense(512))

model.add(layers.LeakyReLU(alph
a=0.2))

model.add(layers.Dense(1024))
```



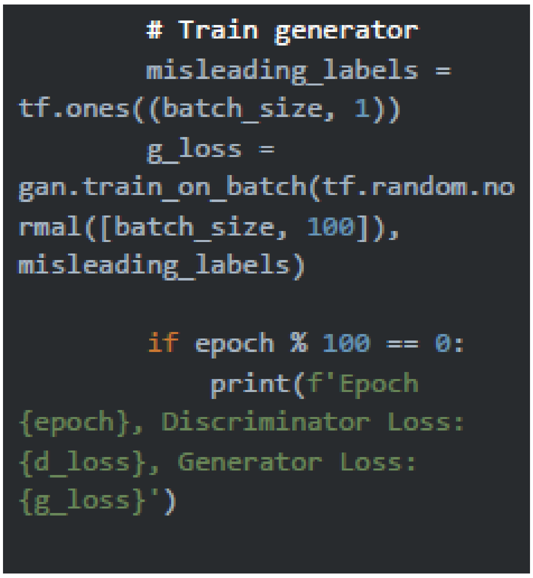
```
model.add(layers.LeakyReLU(alph
a=0.2))

model.add(layers.Dense(data_sha
pe, activation='tanh'))
    return model

# Discriminator model
def build_discriminator():
    model =
tf.keras.Sequential()
```



```
model.add(layers.Dense(1024,
input_shape=(data_shape,)))

model.add(layers.LeakyReLU(alph
a=0.2))

model.add(layers.Dense(512))

model.add(layers.LeakyReLU(alph
a=0.2))

model.add(layers.Dense(256))

model.add(layers.LeakyReLU(alph
a=0.2))
    model.add(layers.Dense(1,
activation='sigmoid'))
    return model

# Training the GAN
def train_gan(generator,
discriminator, epochs,
batch_size, data):
    for epoch in range(epochs):
        # Train discriminator
        real_data =
```



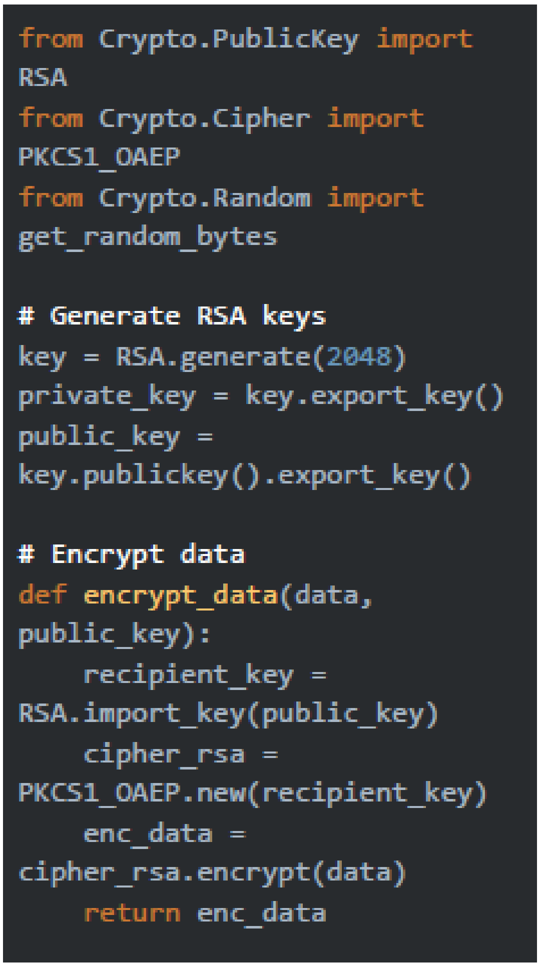
```
data.sample(batch_size)
        synthetic_data =
generator.predict(tf.random.nor
mal([batch_size, 100]))
        combined_data =
tf.concat([real_data,
synthetic_data], axis=0)
        labels =
tf.concat([tf.ones((batch_size,
1)), tf.zeros((batch_size,
1))], axis=0)
        d_loss =
discriminator.train_on_batch(co
mbined_data, labels)
```



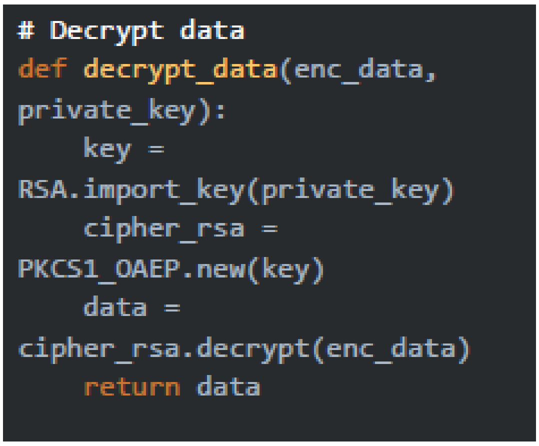
```
        # Train generator
        misleading_labels =
tf.ones((batch_size, 1))
        g_loss =
gan.train_on_batch(tf.random.no
rmal([batch_size, 100]),
misleading_labels)

        if epoch % 100 == 0:
            print(f'Epoch
{epoch}, Discriminator Loss:
{d_loss}, Generator Loss:
{g_loss}')
```

Private and Public Key

Encryption RSA encryption keys are generated by providing a pair of public and private keys. These keys are essential for securing data through encryption and decryption processes.



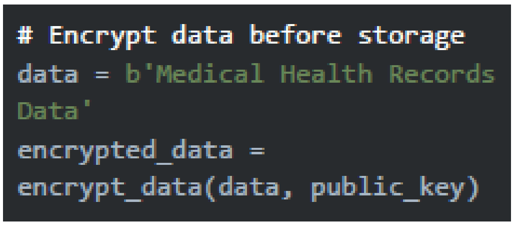
```
from Crypto.PublicKey import
RSA
from Crypto.Cipher import
PKCS1_OAEP
from Crypto.Random import
get_random_bytes

# Generate RSA keys
key = RSA.generate(2048)
private_key = key.export_key()
public_key =
key.publickey().export_key()

# Encrypt data
def encrypt_data(data,
public_key):
    recipient_key =
RSA.import_key(public_key)
    cipher_rsa =
PKCS1_OAEP.new(recipient_key)
    enc_data =
cipher_rsa.encrypt(data)
    return enc_data
```



```
# Decrypt data
def decrypt_data(enc_data,
private_key):
    key =
RSA.import_key(private_key)
    cipher_rsa =
PKCS1_OAEP.new(key)
    data =
cipher_rsa.decrypt(enc_data)
    return data
```

Data Encryption at Source

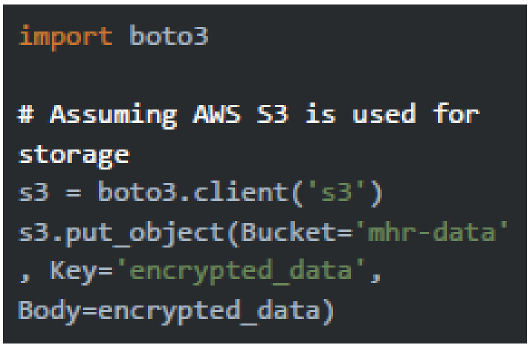
The synthetic data is encrypted using the RSA public key, ensuring that it can only be decrypted by the corresponding private key, protecting the data from unauthorized access.



```
# Encrypt data before storage
data = b'Medical Health Records
Data'
encrypted_data =
encrypt_data(data, public_key)
```

Data Storage Layer Security

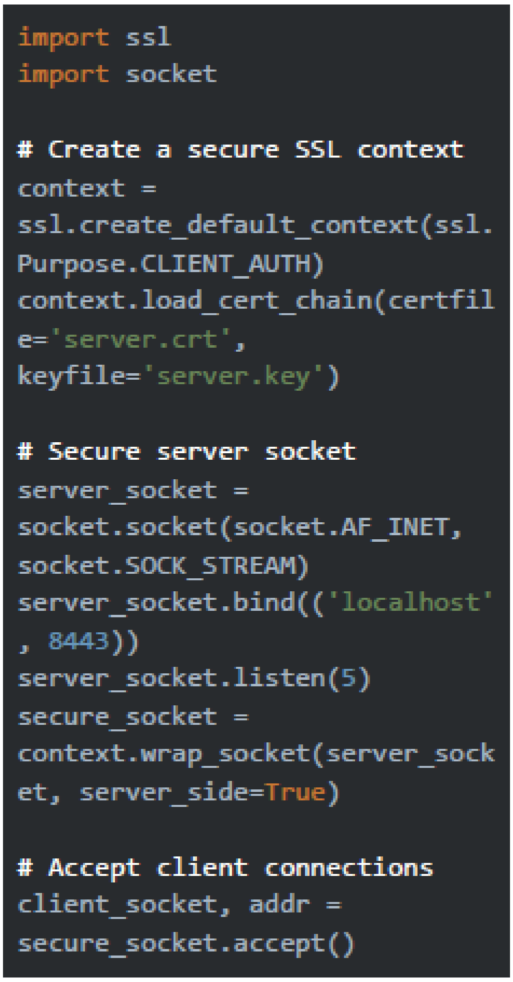
The encrypted data is securely stored in an AWS S3 bucket, leveraging the cloud's robust security measures.



```
import boto3

# Assuming AWS S3 is used for
storage
s3 = boto3.client('s3')
s3.put_object(Bucket='mhr-data'
, Key='encrypted_data',
Body=encrypted_data)
```

Client-Server Communication

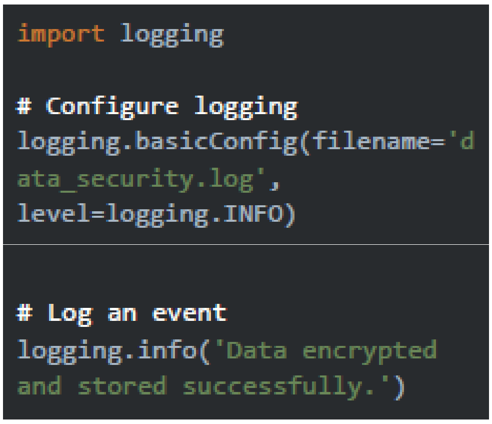
An SSL context is established for secure client-server communication. A secure server socket listens for and accepts client connections, ensuring encrypted data transmission.



```
import ssl
import socket

# Create a secure SSL context
context =
ssl.create_default_context(ssl.
Purpose.CLIENT_AUTH)
context.load_cert_chain(certfil
e='server.crt',
keyfile='server.key')

# Secure server socket
server_socket =
socket.socket(socket.AF_INET,
socket.SOCK_STREAM)
server_socket.bind(('localhost'
, 8443))
server_socket.listen(5)
secure_socket =
context.wrap_socket(server_sock
et, server_side=True)

# Accept client connections
client_socket, addr =
secure_socket.accept()
```

Monitoring and Logging

Logging mechanisms are implemented to track data encryption, storage, and access activities, maintaining detailed logs for monitoring and auditing purposes.

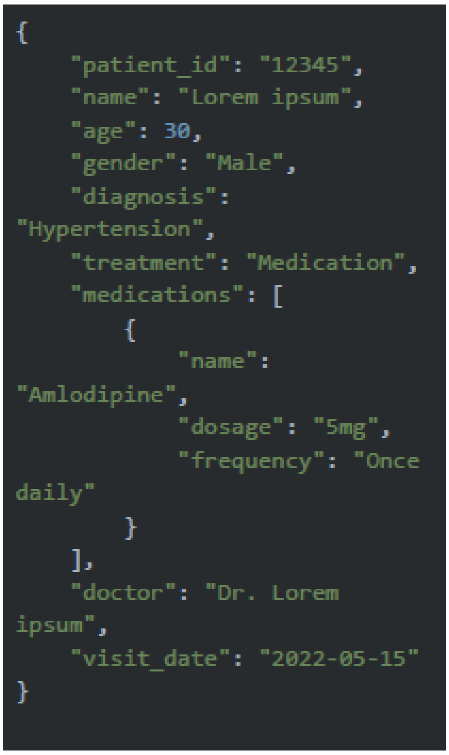


```
import logging

# Configure logging
logging.basicConfig(filename='d
ata_security.log',
level=logging.INFO)


# Log an event
logging.info('Data encrypted
and stored successfully.')
```

Outputs

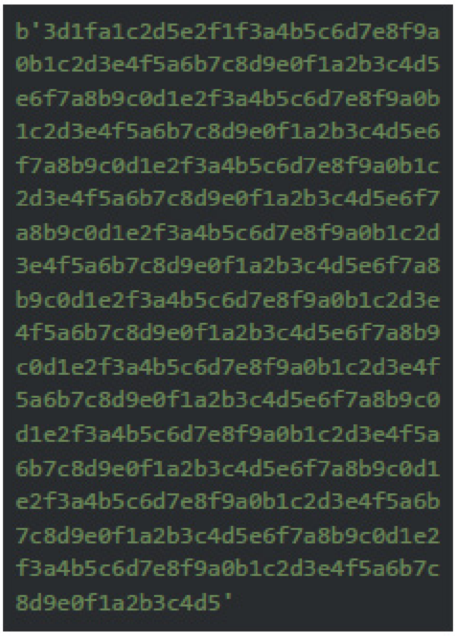
Text Output

A sample input MHR record that has been generated from a medical application is as below:



```
{
    "patient_id": "12345",
    "name": "Lorem ipsum",
    "age": 30,
    "gender": "Male",
    "diagnosis":
"Hypertension",
    "treatment": "Medication",
    "medications": [
        {
            "name":
"Amlodipine",
            "dosage": "5mg",
            "frequency": "Once
daily"
        }
    ],
    "doctor": "Dr. Lorem
ipsum",
    "visit_date": "2022-05-15"
}
```

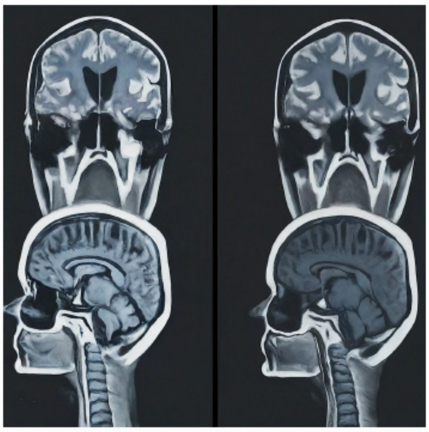
Upon executing the scripts from the 5.1 trigger script, this sample input MHR record is executed, and the Encrypted Data (Hexadecimal Representation for Readability) is as below:



```
b'3d1fa1c2d5e2f1f3a4b5c6d7e8f9a
0b1c2d3e4f5a6b7c8d9e0f1a2b3c4d5
e6f7a8b9c0d1e2f3a4b5c6d7e8f9a0b
1c2d3e4f5a6b7c8d9e0f1a2b3c4d5e6
f7a8b9c0d1e2f3a4b5c6d7e8f9a0b1c
2d3e4f5a6b7c8d9e0f1a2b3c4d5e6f7
a8b9c0d1e2f3a4b5c6d7e8f9a0b1c2d
3e4f5a6b7c8d9e0f1a2b3c4d5e6f7a8
b9c0d1e2f3a4b5c6d7e8f9a0b1c2d3e
4f5a6b7c8d9e0f1a2b3c4d5e6f7a8b9
c0d1e2f3a4b5c6d7e8f9a0b1c2d3e4f
5a6b7c8d9e0f1a2b3c4d5e6f7a8b9c0
d1e2f3a4b5c6d7e8f9a0b1c2d3e4f5a
6b7c8d9e0f1a2b3c4d5e6f7a8b9c0d1
e2f3a4b5c6d7e8f9a0b1c2d3e4f5a6b
7c8d9e0f1a2b3c4d5e6f7a8b9c0d1e2
f3a4b5c6d7e8f9a0b1c2d3e4f5a6b7c
8d9e0f1a2b3c4d5'
```

Image Output

The below image is an example of GAN-based data masking. The left partition of the image is an original X-RAY, whereas the right partition is a masked image.



Conclusion

This white paper presents a strong solution to the enterprise data security challenges in the healthcare sector with advanced data masking techniques using Data Science and GAN. With exponentially growing digitalization of healthcare records, the need for protection of sensitive patient information becomes paramount. Our proposed solution harnesses GAN to create synthetic data, retaining utility like original data and ensuring privacy and compliance with stringent regulations. Data masking, encryption, secure data transmission, protection against cyber-attacks, and preserving data integrity are how healthcare organizations can keep patients' information safe and thus sustain the trust in the institutions.

In the applied landscape of present healthcare data security, there are millions of challenges to surmount: the complexity of healthcare data, regulatory compliance, insider threats, and dynamically changing cyber threats. Handling these challenges in this approach will ensure scalable and efficient data masking techniques for data privacy, improve the compliance situation, and warrant better data utility in all non-production environments. Added to that, the integration of the RSA encryption layer ensures a guarantee that only the intended recipient would have access to the encrypted data and view it plain, thus further strengthening this security framework.

Future recommendations include the continuous improvement of AI and machine learning models underlying data masking techniques for more accuracy and efficiency. The continuous updating of encryption algorithms and security protocols will help the system to remain fit against new forms of threats. Moreover, security awareness and training among healthcare organizations' staff will help avoid insider threats. Continuous monitoring and auditing of data access and usage will help ensure data integrity and compliance with regulations. Advanced technologies and practices can help healthcare organizations increase the security of their data, aspects of patient privacy, and improve quality of care in general.

References

1. Lim LJ, Tison GH, Delling FN (2020) Artificial intelligence in cardiovascular imaging. Methodist DeBakey Cardiovascular Journal 16: 138-145.
2. Nguyen CT, Liu Y, Du H, Hoang DT, Niyato D, et al. (2021) Generative AI-enabled blockchain networks Fundamentals applications and case study. IEEE Network 35: 49-55.
3. Luo S, Chen J (2022) A comparative study of machine learning and deep learning techniques for fake news detection.

Information 13: 576.

4. Guzman AL, Lewis SC (2019) Artificial intelligence and communication A human-machine communication research agenda. New Media & Society 22: 70-86.
5. McKinsey & Company (2022) Generative AI and the future of work in America https://www.mckinsey.com/featured-insights/future-of-work/generative-ai-and-the-future-of-work-in-america.
6. Kooli C, Al Muftah HA (2022) Artificial intelligence in healthcare a comprehensive review of its ethical concerns. AI & SOCIETY 1: 121-131.
7. Landi H (2022) Healthcare data breaches hit all-time high in 2021 impacting 45M people. Health Tech https://www.fiercehealthcare.com/health -tech/healthcare-data-breaches-hit-all-ti me-high-2021-impacting-45m-people.
8. Nguyen CT, Y Liu HDu, Hoang DT, Niyato D, Nguyen DN, et al. (2021) Generative AI-enabled Blockchain Networks: Fundamentals Applications, and Case Study IEEE Network 35: 49-55.