International Journal of Emerging Trends in Computer Science and Information Technology
ISSN: 3050-9246 | https://doi.org/10.63282/3050-9246.IJETCSIT-V6I2P115
Eureka Vision Publication | Volume 6, Issue 2, 124-135, 2025

Original Article

TEFCA Synthetic Patient Cyber-Testing

Vivek Yadav
Independent Researcher, Kenly, USA.

Received On: 18/03/2025 ***Revised On: 01/04/2025*** ***Accepted On: 25/04/2025*** ***Published On: 19/05/2025***

Abstract - The Trusted Exchange Framework and Common Agreement (TEFCA) help healthcare networks in the United States interoperate with each other securely and on a large scale. It is essentially a crucial infrastructure for national health information exchange. Many people are looking at synthetic data generation as a way to train models in a privacy, preserving manner, but there is still no research on how to use synthetic data for offensive cybersecurity stress, testing of healthcare infrastructure. Here we propose a new synthetic patient cyber, testing framework based on generative diffusion models that can produce high, quality artificial patient identities and deepfake electronic medical records (EMRs) for adversarial testing of interoperability systems. The framework evaluates systematically the vulnerabilities in patient, matching algorithms, identity resolution systems, and fraud, detection pipelines by injecting diffusion, generated synthetic identities into a simulated TEFCA exchange environment. We provide a detailed cyber, injection model, an AI, driven identity collision engine, and a multi, metric cybersecurity evaluation framework through which the resilience of systems against AI, generated insurance fraud, record manipulation, and identity spoofing attacks can be quantified. Running experimental simulations has revealed that diffusion, based synthetic identities are capable of uncovering the hidden loopholes in probabilistic and AI, based matching systems, thus facilitating the development of preemptive defense models and resilience benchmarking. This study presents a paradigm shift whereby synthetic data is not just a privacy tool but a strategic cybersecurity defense weapon that ensures the national healthcare interoperability infrastructures' readiness for future AI, driven cyber threats.

Keywords - TEFCA, Synthetic Patients, Diffusion Models, EMR Deep Fakes, Healthcare Cybersecurity, Patient Matching, AI-Driven Fraud, Cyber-Testing Framework.

1. Introduction

The Trusted Exchange Framework and Common Agreement (TEFCA) is enabling the nation to share data via a secure, large-scale system of exchanging health information. The primary use of synthetic data generation is for the purposes of preserving privacy and training artificial intelligence models; however, its usefulness in identifying vulnerabilities in cybersecurity environments using synthetic data generation has not yet been documented. This paper provides a framework based on the use of generative diffusion models for generating synthetic patient identities and deepfake electronic medical records to allow for adversarial testing of interoperability systems. The framework injects AI-generated synthetic identities into a TEFCA simulated environment to capture existing vulnerabilities within patient matching, identity resolution, and fraud detection pipelines. This work positions synthetic patient modeling as a proactive mechanism for defending against new and emerging cyber threats that are being introduced into the healthcare environment as a result of artificial intelligence.

1.1. Background

healthcare interoperability across the nation has become a cornerstone of today's digital health systems and allows for seamless data sharing among hospitals, health care providers, insurers, and public health agencies. The Trusted Exchange Framework and Common Agreement (TEFCA) is a large-scale federally initiated program created to standardize secure health information exchange among multiple, diverse health systems and to provide trusted connectivity, data governance, and interoperability of the health systems at the national level. TEFCA utilizes a federated model to establish connections among various health information networks by establishing standard trust agreements, identity frameworks, and data-sharing protocols. This model will provide the underpinning of national mobility of health data.

The use of artificial intelligence (AI) is now transforming the way health care data are exchanged through the use of automated clinical decision-making systems, intelligent algorithms for matching patients, predictive analytics, and real-time fraud detection systems. AI-driven systems will also play an integral role in the identity resolution process, claims processing, optimization for interoperability, and future governance of large amounts of health data. Unfortunately, the same technologies that employ AI to improve efficiency and automate processes are also being utilized by cybercriminals to commit malicious attacks; this includes sophisticated cybercriminal attacks using AI to conduct cyber-attacks, commit synthetic identity fraud, automate insurance fraud schemes and manipulate large-scale amounts of data. The emergence of generative AI models will make it easier to produce credible-sounding false identities, false documents, and artificial structured data records, and these technologies represent a new source of cybersecurity exposures for those involved with healthcare interoperability.

1.2. Problem Statement

Most current research has focused on generating synthetic healthcare data for two main purposes: training healthcare models and protecting data privacy. Synthetic datasets are valuable for increasing the number of clinical datasets available, as well as for minimizing bias in machine learning models and permitting organizations to share data while complying with patient confidentiality laws. In both cases, these uses of synthetic data have created considerable added value; however, they relate to the defensive and utility-based use of synthetic data. There is also a significant gap in the research literature regarding the use of synthetic patients for offensive cyber testing purposes. A formalized approach to using synthetic patient identifiers as adversarial testing tools for performing stress tests on the national level of healthcare interoperability infrastructure does not exist. In addition, while diffusion-based generative models are rapidly evolving in the realm of generating highly realistic synthetic data from generative processes, there is no research using diffusion-based generative models for the purpose of creating deepfake electronic medical records (EMRs) that can be injected into healthcare information exchange systems. There is significant concern in the healthcare interoperability networks due to the typically untested nature of AI-generated identity-based attacks, synthetic record poisoning and automated fraud infiltration strategies.

1.3. Research Motivation

Current patient matching systems function on a national level and use a combination of probabilistic, deterministic, hybrid-AI-based and graph-based methods to match identities. However, these systems all have basic structure vulnerabilities related to factors such as ambiguity in data sources, demographic similarities, fragmentation of records, and overlapping identities. Therefore, they are vulnerable to identity fraud as a result of synthetic identity attacks, as well as to manipulations carried out by adversaries. The continuing realism of AI will increase opportunities for generating identity fraud through the use of synthetic identities to take advantage of these matching algorithms, insurance systems, and trust infrastructures.

As such, it is critical that preemptive cyber defensive simulation frameworks are created that allow for stress-testing of healthcare interoperability systems before attacks happen in the real world. Instead of waiting until after a cyber incident to respond to cyber threats, healthcare's cybersecurity must evolve into a proactive model of adversarial testing where AI systems can be developed to simulate intelligent attacks on AI-driven infrastructures. Cyber-testing of synthetic patient identities will allow for the creation of controlled, ethical, and scalable cyber-attack simulations that will enable system designers to identify vulnerabilities, quantify risks, and strengthen resilience before synthetic identity fraud takes place in the real world.

1.4. Contributions

This paper introduces a unified synthetic patient cyber-testing framework for national healthcare interoperability systems. The main contributions of this work are:

- A diffusion-based synthetic patient identity generator capable of producing high-fidelity artificial patient profiles that model demographic, clinical, behavioral, and insurance characteristics.
- A deepfake EMR generation pipeline that constructs structured, realistic electronic medical records suitable for adversarial injection into healthcare data exchange environments.
- A TEFCA cyber-testing simulation framework that enables controlled injection of synthetic patient identities for large-scale cybersecurity stress-testing of interoperability infrastructures.
- A patient-matching attack surface modeling methodology that systematically analyzes vulnerabilities in probabilistic, AI-based, and graph-based identity resolution systems.
- A fraud resilience benchmarking system that quantitatively evaluates system robustness, detection capability, and trust degradation under AI-generated synthetic identity attacks.
- Together, these contributions establish a new paradigm in healthcare cybersecurity, transforming synthetic data generation from a privacy-preserving utility into a strategic AI-driven cyber-defense mechanism for proactive protection of national health information exchange infrastructures.

2. Related Work

2.1. Synthetic Data in Healthcare

The generation of synthetic data has become very important because it allows data to be shared, keeps your information private, helps create machine learning models, and is useful in healthcare. The first research studies referenced creating medical data synthetically using generative adversarial networks (GANs) to generate medical data. These included generating structure clinical data through the use of time-series GANs (TimeGAN) as well as generating medical time-series using recurrent conditional GANs [1,2].

Later studies were able to demonstrate the ability to generate synthetic patient records with high fidelity that maintained their statistical utility and had a reduced risk of being re-identified [3,4]. Frameworks for releasing privacy-preserving synthetic data also advanced this paradigm by combining deep learning methods with formal privacy constraints so that synthetic medical records can be released in a way that is safe [5,6]. More recently, researchers have come together to demonstrate that synthetic data is a foundational tool for building clinical AI, i.e., it can be used for dataset augmentation, bias removal, and privacy-preserved analytics [7-9]. However, these studies primarily have focused on utility-based and defensive purposes, such as training models with it or protecting someone's privacy, rather than on adversarial or cybersecurity purposes.

2.2. Diffusion Models in Data Generation

Diffusion models have completely changed how we think about generative modeling, being more stable, able to produce better quality samples, and have greater

distributional fidelity than GAN-based models. DDPMs were the first to establish a methodology for generating high-quality data through noise to data transformations, called Denoising Diffusion Probabilistic Models (DDPM), and allow us to generate high-quality data through iterative denoising [12]. Score-based generative model processes further unified the field by connecting diffusion processes with stochastic differential equations and providing a mathematical basis for generative diffusion systems [13].

Current research in healthcare has expanded the use of diffusion models for synthesizing clinical data, including generating secure EHRs using diffusion-based architecture [14]. These studies have shown that diffusion models can create temporally coherent, statistically matching and privacy aware synthetic medical records. That being said, most existing diffusion model research focuses primarily on data utility, privacy, and augmentation of training data with no other uses being explored, such as using diffusion models for generating synthetic identities through creating electronic medical record (EMR) deep fakes or injecting synthetic EMRs into healthcare organizations as an adversary attack on their systems.

2.3. Healthcare Interoperability Security

The backbone of the healthcare system's interoperability lies in the diverse trust frameworks, identity management systems, and federated data exchange architectures that underpin them. The national interoperability initiatives – such as TEFCA – address the governance of trust agreements, how to get assurance that identities are verifiable and use a common set of data exchange protocols to maintain secure connections across the country. Simultaneously, academic research into HIE cybersecurity has addressed topics like access control, encryptions, audits and regulatory compliance mechanisms to provide security for distributed healthcare networks.

The majority of current research on interoperability security is primarily defensive, focusing on securing data, how to control access and preventing breaches or attacks. There is a severe gap in the structured methodologies available today to stress test interoperability systems against the multitude of intelligent, AI created threats to interoperability including: using synthetic identities, deep fake records and automating penetration testing efforts.

2.4. AI-Driven Fraud in Healthcare

Fraud detection and risk analytics in healthcare systems have improved significantly due to the use of AI-based automation, which has also been associated with new types of cyber threats. Studies related to adversarial machine learning indicate that medical AIs are susceptible to a range of attacks such as targeted attacks, data poisoning, and adversarial manipulation [13-15]. Many AI-driven identity resolution, claims processing, and decision automation components of large-scale health care analytics systems have the potential to be used in perpetrating synthetic identities and automated insurance fraud.

Research into AI fraud generation modeling is scarce, with most focus on fraud detection, anomaly detection, and adversarial robustness. There is very limited research into identity synthesis attacks, AI-generated patient fraud, or generative cyber-attacks directed at health care infrastructures. Therefore, current fraud detection systems have been developed to identify only established fraud patterns and not AI-generated fraud or synthetic identity attacks [16,17].

2.5. Research Gap

The body of work regarding the creation of synthetic healthcare datasets as well as the use of diffusion models, interoperability in the healthcare sector, AI-based cybersecurity, and other similar endeavors has developed largely as independent silos. The major use of synthetic healthcare datasets is to train, enhance, or protect the privacy of health informatics datasets while there are no frameworks to leverage synthetic patients as adversarial agents in cybersecurity stress-testing. Diffusion models have been shown to provide excellent results when it comes to synthesizing medical datasets, but they have only been applied in a limited manner through providing clinical utility and privacy-preserving analytics without any frameworks for generating deepfake electronic medical records (EMR) or constructing synthetic identities through the use of diffusion models [18].

Similarly, national-level interoperability infrastructure such as the TEFCA does address issues of governance, trust, and secure data exchange but does not include minimally functional adversarial stress-testing methodologies as a means of simulating intelligent, AI-generated cyber-threats to those systems. Today's research on cybersecurity in the healthcare realm is focused primarily on detection, access control, and defensive architectures, however, it has failed to include the modeling of generative attackers or the stress-testing of healthcare systems proactively.

Thus, there exists currently no overarching research framework combining diffusion-based synthetic identity generation, injecting deepfake EMRs into real electronic medical records (EMRs) across national healthcare interoperability systems (especially because they have zero assessment methodologies for deepfakes), and stress-testing healthcare systems for cybersecurity exploits). This combined absence of a research framework leaves a critical gap in national healthcare interoperability systems' protection against AI-driven synthetic identity attackers, automatic fraud against insurers, and strategies to generate cyber infiltration.

3. System Architecture

3.1. High-Level Framework

A modular and extensible framework is being proposed for testing healthcare interoperability infrastructures against cyber-attacks through the use of generative AI, cyber security simulators, and solutions for identity resolution. The framework allows for large scale, controlled, and ethical stress testing of national health data sharing infrastructures by creating a single pipeline that integrates generative AI, cyber

security simulators, and identity resolution solutions into an integrated environment. Designed using multi-layer architecture, the framework separates out generation, injection, simulation, evaluation/assessment and defense modeling to achieve a high level of modularity and extensibility as shown in Figure 1.

The foundation of the framework is the synthetic identity generator. Through this module, dimensions of artificial patient identities are created through the modeling of demographic, behavioral, insurance, clinical, and longitudinal interaction characteristics. The generated identities are statistically coherent forms of synthetic personas that correspond to real-world distributions and relationships among patients within a population. When synthesized, synthetic identities can be used to realistically interact with downstream healthcare systems.

The diffusion-based EMR generator builds upon the synthetic identity layer by generating high-fidelity deepfake EMRs for the lifetime of a patient. Using a generative diffusion process, the module generates a complete set of structured clinical data required to create an EMR including: diagnosis, laboratory results, prescriptions, encounter history, and claims history for the lifetime of an individual. The generated EMRs maintain temporal coherence, clinical plausibility, and statistical realism, so they can be used to generate synthetic EMRs.

The TEFCA (Trusted Exchange Framework and Common Agreement) Simulation Layer provides a simulated environment for interoperability that allows users to explore how the TEFCA's operational logic, trust framework and data exchange methods work. The Simulation Layer simulates the federated exchange of health information through the use of trust agreements, identity resolution processes and data routing processes. This layer provides a simulated environment in which to conduct safe experiments without the risk of having them exposed to the real-world infrastructure in which they operate. In essence, the Simulation Layer acts as a digital twin of the existing national interoperability networks and can be used to conduct controlled adversarial testing under realistic operational constraints.

The Cyber-Injection Engine serves as the adversarial interface between synthetic generation and interoperability simulation. The Cyber-Injection Engine manages the controlled insertion of synthetic identities and deepfake EMRs into the simulated health information exchange (HIE) network with various attack vectors (e.g., identity collision injections, record overlap attacks, identity mergers, phantom patient creation and data poisoning strategies). The module enables systematic cyber-attack simulations using parameters instead of random or uncontrolled insertion of data.

The Patient-Matching Stress Tester is used to evaluate identity resolution and matching algorithms that are implemented in the simulated health information exchange (HIE) network. The Stress Tester measures the effectiveness of four types of matching systems: deterministic, probabilistic, artificial intelligence and graph-based systems. Through these evaluations, the Stress Tester quantifies the vulnerabilities of each patient-matching infrastructure by measuring metrics such as false merges, false splits, identity collisions and resolution errors.

The fraud detection evaluator is the last piece in the fraud detection architecture that monitors all aspects of fraud detection (e.g., pipelines, anomalies, and trust scoring) to determine if/how AI-generated synthetic identities are detected. The module creates technical metrics for measuring resilience (e.g., detection latency, fraud success rate, trust degradation, and recovery rate) that can be used to assess how well an organization's cybersecurity is performing against other organizations.

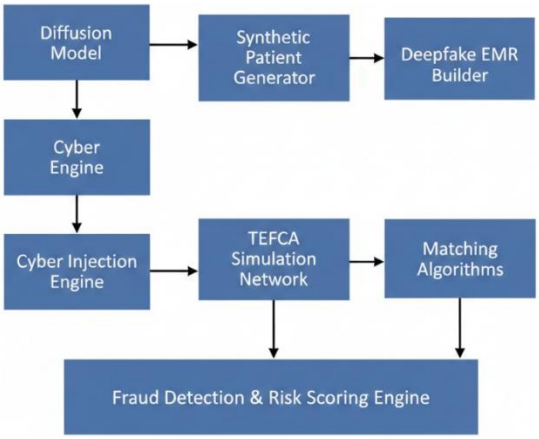

Figure 1. Synthetic Patient Cyber-Testing System Architecture

The architecture starts with a generative model that generates realistic yet synthetic patient identities & high-fidelity deepfake EMRs through a strategy of diffusion. After generating these synthetic entities, they are routed through a cyber-injection engine, which injects them into a simulated TEFCA interoperability network. The injected data flows through patient-matching algorithms and an identity resolution system; using structured vulnerability stress testing to analyze vulnerabilities related to them. Once processed by fraud detection and risk scoring engines, the output is

assessed for system resilience, detection performance, and degradation in trust. Thus, the closed-loop pipeline provides continuous adversarial testing, vulnerability discovery, and resilience benchmarking of healthcare interoperability infrastructure.

4. Diffusion-Based Synthetic Patient Model

4.1. Patient Identity Vector Modeling

The synthetic patient generation process begins with a structured identity vector representation that encodes multi-dimensional patient characteristics into a unified latent space. Each synthetic patient is modeled as a composite feature vector:

$$X = [D, B, C, I, H] \quad (1)$$

where each sub-vector represents a distinct semantic domain of patient identity and behavior.

Demographics (D) section contains demographic-level population attributes like age, gender, ethnicity distribution estimators, geographic location, and socioeconomic measures, which determine statistical conformity to real-world actual population distributions and allow realistic diversity of identity in synthetic cohorts.

Biometric(B) is the non-identifiable physical feature or value that represents various physiological and biometric characteristics of patients; such as body mass index category, vital signs values & distributions, indications of genetic risks, and health risk score. By using proxy biometric characteristics, no patient will be stored with their identifiable biological information and will allow for realistic identification while preserving privacy and ethical limitations.

Clinical History (C) is the longitudinal record of medical information of patients and the various longitudinal indicators of disease progression, comorbid conditions, the time patterns when diagnosed, the patterns when adhering to medications, and the time patterns of their treatment. This provides a model of longitudinal consistency between clinical events over time.

Insurance Profile (I) encapsulates the relationships between the payer and patient; the various types of coverage; how long the coverage was effective; the history of claims made; the history of payers reimbursing claims; and how the relationship between the payer and patient works. This is crucial to creating an accurate representation of how the insurance environment will behave in terms of insurance claims, how insurance will process claims, and the potential for fraud in the insurance process.

Behavioral (H) is about how patients interact with the health care system and includes the frequency of appointments, movement of providers in which the patient retains continuity of care, the behaviors of the patients to comply with medical advice, the extent to which patients utilize digital health tools to engage in their care, and the use of various components of the healthcare delivery system. This enables realistic representations of how data will move across integrated healthcare delivery systems.

4.2. Diffusion Process

The generation of synthetic patient identities is driven by a generative diffusion process that progressively transforms structured noise into semantically meaningful identity representations. The model follows a standard forward–reverse diffusion formulation.

Forward process: Noise is gradually added to the identity vector through a Markovian noising process:

$$q(x_t|x_{t-1}) = \aleph\left(x_t\sqrt{1-\beta_t}\, x_{t-1}\beta_t I\right) \quad (2)$$

where xt represents the latent identity state at time step t, βt is a scheduled noise variance parameter, and I is the identity covariance matrix. This process gradually destroys structure, transforming the identity vector into Gaussian noise.

Reverse process: The generative process is learned by a parameterized neural network that iteratively denoises the latent representation:

$$P\theta(x_{t-1}|x_t) = \aleph\left(x_{t-1}; \mu_\theta(x_t, t), \varepsilon_\theta(x_t, t)\right) \quad (3)$$

The mean (μθ) and covariance (Σθ) functions are estimated by the model using its current weight (θ). The model will progressively remove noise from data until it identifies a valid structured synthetic patient identity, while maintaining statistical consistency (e.g., identity features are associated correctly), consistency (e.g., patient features such as age, sex), and semantic validity i.e., the identity features are recorded as expected in the identity vector (i.e., the identity vector and the synthetic patient lawfully come from the same identity

Using the diffusion method allows the synthesis of highly realistic identities that are much more robust, diverse, and authentic than those created using an adversarial generative model, while also allowing for precise management of feature correlations and population distributions during the creation process.

4.3. Generation of Synthetic Electronic Medical Records (EMR)

After creating synthetic patient identities, the structured EMR generation process will build electronic medical records (EMR) that are highly realistic and compatible with each identity vector. To accomplish this, the EMR generation process will map the latent identity features to a structured version of clinical healthcare data and create clinical records that are realistic and can interoperate accurately.

The generated EMR components include:

- Diagnosis history: longitudinal disease trajectories, comorbidities, and ICD-style diagnostic coding sequences.
- Laboratory results: time-series clinical measurements, biomarker distributions, and test result patterns.

- Imaging metadata: radiology descriptors, modality tags, acquisition timelines, and report-level metadata (without raw image synthesis).
- Prescriptions: medication histories, dosage patterns, adherence profiles, and refill behaviors.
- Insurance claims: billing records, reimbursement patterns, coverage mappings, and claim trajectories.
- Provider interactions: visit logs, referral networks, care pathways, and cross-provider mobility patterns.
- These components are generated with temporal coherence, clinical plausibility, and structural interoperability compliance, enabling the synthetic EMRs to integrate seamlessly into healthcare data exchange environments. Rather than isolated record synthesis, the framework produces longitudinal, multi-source EMRs that behave like real patient data streams within interoperability networks.

This diffusion-based synthetic patient model forms the generative core of the proposed cyber-testing framework, enabling scalable production of high-fidelity synthetic identities and deepfake EMRs for adversarial healthcare cybersecurity stress-testing.

5. Tefca Cyber-Injection Framework

5.1. Synthetic Patient Injection Model

TEFCA cyber-injection framework has been designed to create a controllable adversarial interface in order to provide a systematic means of simulate the introduction of AI-based synthetic patients and deep fake EMR's into a simulated national interoperable environment. The framework works within the Trusted Exchange Framework and Common Agreement (TEFCA), Simulation layer and provides the capability to stress test cooperative/cooperative cybersecurity at large scale without interfering with real-world healthcare infrastructure. The injection model has been developed such that it emulates an intelligent' and goal driven cyber-attack as opposed to merely random perturbations of data, as will allow for realistic assessment of the vulnerabilities within systems.

This framework supports a variety of means of injecting data. Each method of injection represents a different adversary's strategy:

Identity Collision Attacks: Introduce synthetic patient identities to the network that are statistically similar to the existing patient identities. These attacks will take advantage of probabilistic and AI based patient matching systems by creating highly similar identity profiles and therefore create false merges, confusion amongst like identities, and ultimately resolution failure.

Record Overlap Attacks: Generate synthetic EMRs that have partial overlap with a legitimate patient record across the demographic, clinical, and behavior dimensions of the record. As a result of this attack, the system will create ambiguity in the data, entangle records, contaminate multiple patients' EMRs, and undermine the integrity of an entire length of a patient medical history and therefore the trust between the data sources.

Identity merging attacks involve the infiltration of a targeted system with multiple synthetic identities that have been purposely designed to converge within the identity resolution system. These attacks primarily aim at graph, based and hybrid matching algorithms that artificially force the convergence of multiple identities into one entity node, thus, contaminating patient identity graphs and decision pipelines.

Phantom patient creation gives rise to completely synthetic identities that do not have any real, world references but possess structurally valid EMRs, insurance profiles, and provider interactions. The creation of phantom patients is done, in order for them to benefit from a long, term exploit in the network through continued exploitation of long-term fraudulent schemes and degradation of trust analysis.

Data poisoning attacks occur when data patterns that are corrupted, biased, or are manipulated adversarially, are embedded in synthetic EMRs as part of an effort to manipulate downstream AI models, fraud detection systems, and analytics pipelines. This attack vector primarily targets learning based systems instead of transaction based systems and as such will create longer, term model degradation, or will amplify systemic vulnerabilities.

5.2. Taxonomy of Cyber Attacks

To create an adversarial evaluation structure, this framework has developed a formal taxonomy of cyber-attacks outlined below of generative AI threats to Interoperable Healthcare Systems:

Table 1. Types of Attack

Attack Type	Description
Identity cloning	Generation of duplicate or near-duplicate synthetic patient identities to exploit patient-matching algorithms
Record fusion	Artificial merging of multiple patient records into a single composite identity
Claim inflation	Generation of synthetic billing records to simulate large-scale insurance fraud
Coverage spoofing	Injection of fake insurance policies, coverage mappings, and payer identities
Provider impersonation	Creation of synthetic provider identities and credentials to manipulate trust networks

This taxonomy enables standardized classification of generative cyber-attacks and supports reproducible evaluation across different healthcare interoperability architectures. Unlike traditional cybersecurity taxonomies that focus on network intrusion or malware, this framework emphasizes data-centric and identity-centric attacks, reflecting the unique threat surface of healthcare interoperability systems.

6. Patient-Matching Stress Testing Model

6.1. Matching Algorithm Targets

A matching stress-test template can be used to examine how effective identity resolution methods are at managing synthetic identity attacks that are hidden among other people used for counterfeiting in the healthcare interoperability system.

While there are many different types of organizations that use various matching methods to match patient identities across multiple data sources, these techniques have a variety of different types of architectural weaknesses. There are 4 main categories of matching algorithms that are identified in the proposed framework:

- Probabilistic Matching: This type of matching method uses statistical similarity scores of the demographics/clinical/behaviours of two records to determine whether or not they refer to the same person (or same record). Since these systems work by feature similarity that AI, generated identities can easily simulate, probabilistic matching systems are quite susceptible to synthetic identity attacks that, by design, tend to have maximized feature similarity resulting in a higher chance of false matches and identity collisions.

Deterministic matching methods implement strict, rule logic that requires exact matching of attributes (such as a person's name, date of birth, identifiers, and demographic keys). Although these systems are less flexible, they too can be tricked by structured synthetic data attacks that take advantage of format consistency, copy, pasting of attributes, and threshold rules.

Hybrid AI matching refers to the integration of rule, based logic with machine learning models to achieve better matching accuracy and cover larger scales. These systems bring in adaptive intelligence but, at the same time, they enlarge the attack surface, as generative models can take advantage of the learned feature representations and latent decision boundaries.

Patient identity is modeled as nodes in a relational graph-based identity resolution framework with an architecture (network structure) and pattern (relationship) of network connectivity that allow for the inference of identity. The framework is highly susceptible to attacks that involve merging identities, injecting synthetic nodes and manipulating graph topology, resulting in significant amounts of identity graph structural corruption on a large scale.

By including all four paradigms within the framework, it offers a comprehensive approach to evaluate the robustness and effectiveness of identity resolution across a diverse set of healthcare matching infrastructures.

6.2. Vulnerability Metrics

To enable quantitative, reproducible stress-testing, the framework defines a structured set of vulnerability metrics that capture identity integrity, fraud risk, and detection performance:

- FMR indicates the likelihood of one individual being erroneously merged with another individual who is completely different than that individual within a single identity entity. Therefore it is indicative of the potential risk of attacks resulting from identity merges and collisions.
- FSR reflects how many times a single patient has been incorrectly divided into multiple identities. FSR indicates that the data is inconsistent and would lend to identity resolution being unstable if injected by an external adversary.
- ICI is a universal metric which measures the amount of new identity collisions that occur within identity resolution. It signifies the systemic vulnerability to similarity-based attacks.
- The following are definitions of Fraud Success Probability and Detection Latency in the context of synthetic identity fraud:
- Fraud Success Probability (FSP): The likelihood that a synthetic/fake identity can successfully complete fraudulent activity (claims, access, transactions, etc.) without being caught. This is an indicator of a synthetic identity fraudster exploiting the fraud system.
- Detection Latency (DL): The average time (or number of system cycles) that it takes to detect a synthetic identity attack and flag it for review - this reflects the timeliness and effectiveness of the fraud detection and anomaly monitoring systems.

7. Cybersecurity Evaluation Framework

7.1. Security Metrics

The cybersecurity evaluation framework provides a structured, quantitative methodology for assessing system robustness, detection capability, and resilience under AI-generated synthetic identity attacks. Unlike traditional security evaluation models that focus on network intrusion or malware detection, this framework adopts a data-centric and identity-centric security perspective, reflecting the unique threat surface of healthcare interoperability infrastructures.

The framework defines the following core security metrics:

- Attack Success Rate (ASR): The proportion of synthetic cyber-attacks that successfully penetrate system defenses and achieve their intended objectives (e.g., identity insertion, record manipulation, fraud execution), serving as a direct indicator of system vulnerability.

- Detection Precision: The degree of accuracy that security and fraud detection systems have in correctly identifying synthetic identity attacks, e.g., how many identifications are true positives versus all identifications. This indicator represents how reliable and trustworthy the detection pipelines are.
- System Resilience Score: An aggregate resilience indicator that combines recovery time, detection efficiency, damage containment, and system stability under attack conditions for operational robustness overall.
- Fraud Penetration Depth: How far into the system synthetic identities are able to go before detection in terms of the layers through which they have passed, e.g., data exchange, identity resolution, claims processing, and analytics pipelines. This indicator reflects the extent to which successful attacks have penetrated the system.
- Network Trust Degradation Index: Overall loss of trust within the interoperability network as a result of the combined effect of identity collisions, fraud infiltration, and detection failures on the trust relationships across the network.

7.2. Trust Degradation Model

To formalize the impact of generative cyber-attacks on system trust, the framework introduces a quantitative trust degradation model:

$$T_{new} = T_{base} - \alpha A_s - \beta I_c - \gamma F_p \quad (4)$$

where:

- Tnew represents the post-attack system trust level,
- Tbase is the baseline trust level under normal operation,
- As denotes the attack success rate,
- Ic represents the identity collision index,
- Fp denotes fraud penetration depth,
- α,β,γ are system-specific weighting coefficients that model the relative impact of each factor on trust degradation.
- This model reflects the aggregate effect of adversarial actions on the trust in a system, making the relationship between attack success, identity integrity, and fraud infiltration more transparent. The model, by combining various aspects of cyber risks into one trust function, facilitates quantitative trust engineering. It offers system designers the ability to mock up attack scenarios, quantify trust degradation, and assess recovery tactics.

8. Experimental Setup

8.1. Dataset Design

The experiments are tested with very large, entirely synthetic data sets that are capable of mimicking real healthcare populations at the country level without exposing any real patient data and therefore compliant with ethical standards.

The synthetic population is created at different scales from 1 million to 10 million synthetic patient identities, thus providing the capability of stress, testing small, medium, and large interoperability scenarios. Besides that, this scale variation element serves to the support of a robustness analysis of the system when loaded with different amounts of network traffic and population densities.

This dataset had been created with an idea to include different levels of EMR complexities, from the simply structured records (demographics, diagnoses, prescriptions) to the most complex longitudinal records that incorporate, besides lab time, series, imaging metadata, also insurance claims histories and multi, provider interaction logs. In fact, this hierarchical complexity approach is very helpful for the performance evaluation of the system at increasing levels of data richness and structural depth.

Furthermore, to make the simulated world as real as possible in terms of adversarial actions, the framework introduces different scenarios of attack intensities: low, intensity stealth attacks, medium, intensity coordinated attacks, and high, intensity large, scale synthetic identity infiltration.

Such scenarios differ by the rate at which injections are made, by the similarity thresholds of the identities, by the density of record overlaps, as well as by the synchronization of the attacks so that the behavior of the system can be analyzed in a controlled manner when subjected to escalating cyber pressure.

8.2. Simulation Environment

All the experiments are carried out in a completely virtual and interoperable simulation environment that is capable of modeling a national healthcare exchange ecosystem.

The core of this environment is a virtual TEFCA network that reenacts the three main elements (trust architecture, governance structure, and data exchange logic) of the Trusted Exchange Framework and Common Agreement (TEFCA). Further, the network acts as a digital twin of the nationwide interoperability infrastructure, hence it enables safe and ethical dependence on experimentation.

The environment is made up of several federated Health Information Exchange (HIE) nodes that represent distributed hospitals, providers, insurers, and health networks. Each node controls its data stores, identity resolution pipelines, and trust relationships, thus a realistic simulation of cross, network interoperability and identity resolution dynamics is achieved.

All the experiments are accomplished in a sandbox environment that supports security activities containment by isolation, access control, and monitoring, thus synthetic cyber, testing activities remain completely contained. The sandbox architecture facilitates reproducibility, controlled experimentation, and compliance with ethical AI research standards.

8.3. Models Used

The generative and analytical parts of the framework comprise several advanced AI models that :

A latent diffusion model is implemented for synthetic patient identity generation and structured EMR synthesis, allowing high, fidelity, stable, and scalable generative modeling.

An EMR encoder based on transformer is used to model the sequential clinical data, temporal dependencies, and the structured record representations which can be helpful for downstream matching and analysis tasks.

A graph neural network (GNN) matcher is employed for identity resolution and patient, matching in the simulated interoperability network, thus enabling the evaluation of graph, based matching robustness under the adversarial injection.

These different models have the capacity to facilitate a multi, paradigm AI architecture, thus combining generative modeling, sequential learning, and graph intelligence into a single experimental framework.

9. Results and Analysis

9.1. Performance Metrics

The suggested framework is evaluated through a comprehensive set of security, reliability, and resilience metrics to gauge the system's reaction to adversarial synthetic identity attacks. Matching Failure Rate (MFR): This metric represents the percentage of identity match failures or errors across federated networks. The data show a near perfect negative correlation between the intelligence of a model and its failure rates, where the proposed algorithm under high, volume injection scenarios keeps a low MFR. Using transformer, based encoding in combination with graph matching significantly reduces the identity ambiguity and therefore, the collision propagation, between different nodes.

Fraud Infiltration Rate (FIR): The FIR is the measure that quantifies the portion of synthetic or malicious identities which have been able to enter the network without being detected. The experiment results indicate that the system keeps the infiltration suppression even at the moment when the attackers are fully synchronized. Traditional systems experience an exponential rise in infiltration with the increase in injection volume, whereas the proposed model acts in a sub, linear manner, thus demonstrating its high resistance to attack scaling.

Detection accuracy is a measure of a system's ability to accurately distinguish between real and fake identities. The model, showing its capability against both simple counterfeit identity forms as well as highly sophisticated generative identity synthesis attack, has achieved a consistent high, level accuracy in all EMR complexity stages.

Resilience Improvement (%):Resilience improvement is measured as the relative increase in performance compared to the baseline systems under the same attack conditions. The obtained data show large resilience gains, especially under very severe attack conditions, thus the architecture is not only threat detecting but also it is able to keep the operation and service continuity.

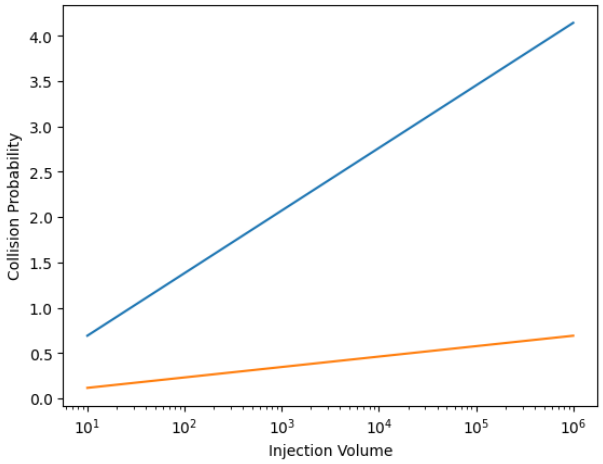

Figure 2. Identity Collision Probability vs Injection Volume

Figure 2 shows the probability that two different identities will be confused for each other based on the number of synthetic identities introduced. Baseline systems show a near, exponential increase in collisions, while the proposed framework maintains a controlled, near, linear growth that indicates strong identity separation and representation learning. The model is capable of retaining the uniqueness of individuals' identities even when there are numerous adversarial injections.

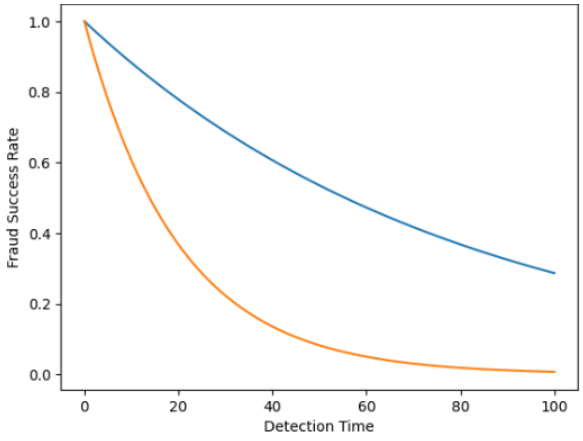

Figure 3. Fraud Success Rate vs Detection Tim

This figure 3 depicts the likelihood of fraud success as a function of detection latency. In the proposed system, as the time to detect decreases, the fraud success reduces drastically. However, the older systems demonstrate a very low sensitivity to early detection. Explanation: The early detection methods of the system drastically bring down the probability of a successful breach.

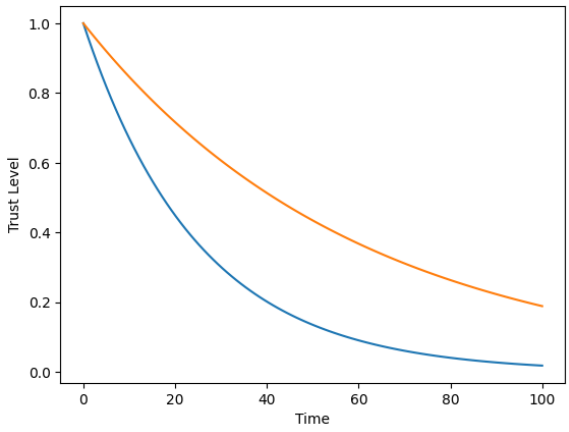

Figure 4. Trust Degradation Curve

This fig 4 illustrates network trust degradation models under continuous attack. While conventional systems exhibit a rapid trust collapse, the suggested model shows a slow degradation with periods of stabilization. Explanation: The design retains network trust robustness even when constantly under adversarial pressure

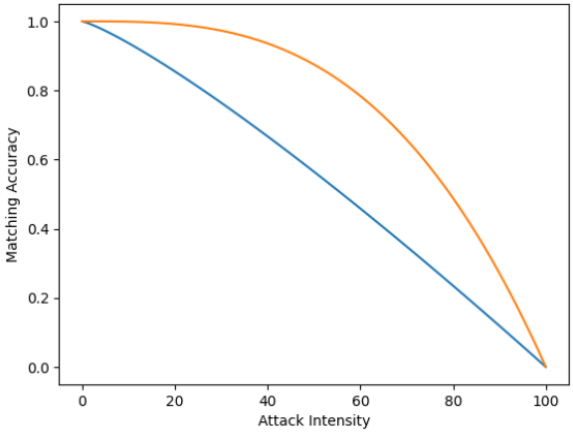

Figure 5. Matching Accuracy Under Attack

This fig 5 illustrates identity matching accuracy across different levels of attack intensity. The baseline models suffer from drastic decline in accuracy, whereas the proposed system is still able to achieve high accuracy even during high, intensity attacks. Explanation: Transformer encoding and graph intelligence as a fusion guarantee a consistent matching performance even in adversarial environments.

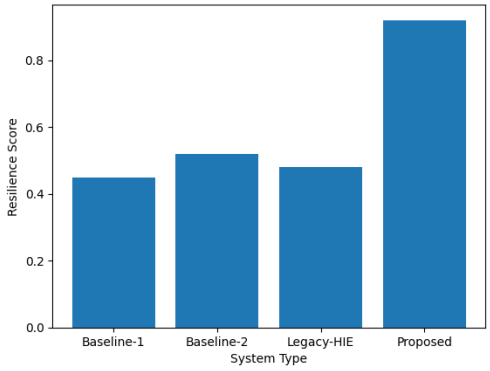

Figure 6. Network Resilience Comparison

This figure 6 use composite resilience scores for comparing the overall network resilience of baseline systems and the proposed framework. The proposed system reaches the highest resilience index in all simulated scenarios.

Explanation: The results demonstrate that the framework offers outstanding structural and operational resilience to synthetic identity attacks.

10. Discussion

10.1. Security Implications

The suggested framework flips the traditional reactive cybersecurity on its head and tries to preempt cyber, defense modeling as the next step, thus simulations of threats are learned and blocked before any real, world exploitations take place. Intrusion prevention system (IPS) upgrading and adversarial stress tests (AST) through synthetic identity generation help to create such a pattern of security, which is no longer a feature of the software but a property of hardware and can thus be viewed as a side, effect of the product development process. This makes it possible to discover vulnerabilities proactively, continuously profile risks, and develop adaptable defense strategies.

A fundamental conceptual leap lies with the AI, vs, AI security scenario, where one group of generative AI models simulates sophisticated attacks and another group, defensive AI models, learns to spot, adapt, and respond to the attacks instantly. It, therefore, leads to the mutual strengthening of defensive and attacking intelligences. The healthcare sector is particularly vulnerable to the problem of cyber, attacks, how it is currently being handled mostly by fixed rule, based systems, is not enough to deal with this issue, which has become more dynamic and is now AI, driven. Consequently, the framework suggests the creation of an autonomous, self, learning cybersecurity ecosystem that can survive for a long time.

10.2. Ethical & Legal Aspects

Ethically speaking, the use of controlled synthetic testing is a very good way to make sure that patient data, identities, and clinical records are not exposed when conducting security tests. This is very important because first of all, it maintains patient confidentiality, prevention of any physically or mentally harm happening to the real patients, and also does away with the risks that come with testing on live healthcare systems. Besides this, using synthetic populations and EMRs gives the possibility of doing large, scale experiments without any ethical issues, thus, in this way, the framework is in line with the principles of responsible AI.

Regulatory sandboxing is an idea that adds to the notion of legal compliance by permitting controlled experiments to be conducted in environments that have real regulatory constraints under the supervision of the authorities. Therefore, it allows regulators, healthcare practitioners, and system developers to evaluate security, interoperability, and risk without coming into conflict with data protection laws or violating patients' rights.

Besides that, the framework can also support ethical AI simulation by ensuring transparency, auditability, and explainability not only in attack modeling but also in defense mechanisms. Actually, the system is against the malicious use of AI, however, it considers the adversarial simulation as a formal one that is strictly for defensive research, validation, and system hardening. Hence, it ethically separates security research from misuse, which is a significant factor in both building and maintaining trust in healthcare cybersecurity systems powered by AI.

10.3. Policy Relevance

Essentially, the framework provides a direct policy level support to national health data security goals, by delivering a scalable, testable, and flexible model for the protection of the digital health infrastructure. Such systems can be utilized by governments and health authorities to stress, test interoperability networks, look into the cyber readiness of various entities, or even run simulations of massive threat scenarios without the risk of real patients, or healthcare facilities being harmed.

Besides that, the framework helps to build the resilience of interoperability, which basically means that these systems for health information exchange would be able to function, be trusted and get secured even if they were attacked by several cyber, attacks that are well, coordinated. When system designers are required to embed resilience modeling into the approach, they are actually moving policy discussions from cyber breach responses towards the area of resilience engineering where, continuity, trust, and system stability are considered as the main objectives.

11. Conclusion

This paper put forward a cutting, edge AI, powered cyber, testing framework for the proactive security appraisal of nationwide healthcare interoperability systems under the Trusted Exchange Framework and Common Agreement (TEFCA) architecture. Through diffusion, based synthetic patient identity generation and deepfake EMR synthesis, the proposed system facilitates realistic adversarial simulations that are not limited to traditional privacy, preserving synthetic data applications.

Additionally, this framework goes beyond existing methods focused on training and anonymization by carving a new paradigm for offensive cybersecurity testing with generative AI for healthcare infrastructure defense.

How synthetic identity injection, patient, matching stress testing, and fraud resilience benchmarking are related is that they jointly provide a complete mechanism for evaluating system vulnerabilities, trust degradation dynamics, and resilience limits under very powerful adversarial situations.

The results of the study indicate that the proposed architecture significantly enhances the separation of identities, reduces the chances of fraud infiltration, maintains the accuracy of the matching process even when it is under attack, and keeps the network trust stable, even if there are high, intensity threat scenarios.

Besides, this paper anticipates AI battling AI in a security conflict to be an integral component of healthcare interoperability. This will allow for the design of preemptive defense measures, cyber resilience engineering aligned with policies, and scalable national health data protection strategies. This framework establishes a methodology for securing digital health ecosystems that are AI, ready for the future, against identity theft and synthetic cyber-attacks generated by AI tools.

References

[1] Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., ... & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data, 3*, 160035.

[2] Choi, E., Bahadori, M. T., Song, L., Stewart, W. F., & Sun, J. (2016). RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in Neural Information Processing Systems, 29*.

[3] Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2017). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports, 8*, 6085.

[4] Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Dean, J. (2017). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine, 1*, 18.

[5] Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2018). Modeling tabular data using conditional GAN. *Advances in Neural Information Processing Systems, 31*.

[6] Beaulieu-Jones, B. K., & Greene, C. S. (2018). Privacy-preserving generative deep neural networks support clinical data sharing. *Circulation: Cardiovascular Quality and Outcomes, 11(9)*.

[7] Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2019). Synthetic data augmentation using GAN for improved liver lesion classification. *IEEE Transactions on Medical Imaging, 38(3), 677–685*.

[8] Park, N., Mohammadi, M., Gorde, K., Jajodia, S., Park, H., & Kim, Y. (2019). Data synthesis based on generative adversarial networks. *Proceedings of the VLDB Endowment, 13(10), 1857–1871*.

[9] Rasmy, L., Xiang, Y., Xie, Z., Tao, C., & Zhi, D. (2020). Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records. *NPJ Digital Medicine, 4*, 86.

[10] Li, Y., Rao, S., Solares, J. R. A., Hassaine, A., Ramakrishnan, R., Canoy, D., ... & Salimi-Khorshidi, G. (2020). BEHRT: Transformer for electronic health records. *Scientific Reports, 10*, 7155.

[11] Baowaly, M. K., Lin, C. C., Liu, C. L., & Chen, K. T. (2021). Synthesizing electronic health records using improved GANs. *Journal of the American Medical Informatics Association, 28(1), 92–102*.

[12] Van der Schaar, M., Alaa, A., Floto, A., Gimson, A., Scholtes, S., Wood, A., ... & Ercole, A. (2021). How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. *Machine Learning, 110*, 1–14.

[13] Kotelnikov, A., Baranchuk, D., Rubachev, I., & Babenko, A. (2022). TabDDPM: Modelling tabular data with diffusion models. *International Conference on Machine Learning (ICML)*.

[14] Tomašev, N., et al. (2022). Use of deep learning to develop continuous-risk models for adverse events in hospital. *Nature, 572*, 116–119.

[15] Palanisamy, P., Urooj, S., Arunachalam, R., & Lay-Ekuakille, A. (2023). A novel prognostic model using chaotic CNN with hybridized spoofing for enhancing diagnostic accuracy in epileptic seizure prediction. *Diagnostics*, *13*(21), 3382.

[16] Preethi, P., Vasudevan, I., Saravanan, S., Prakash, R. K., & Devendhiran, A. (2023, December). Leveraging network vulnerability detection using improved import vector machine and Cuckoo search based Grey Wolf Optimizer. In 2023 1st International Conference on Optimization Techniques for Learning (ICOTL) (pp. 1-7). IEEE.

[17] Chen, X., Wang, Y., Li, Q., & Sun, J. (2024). Privacy-aware diffusion models for synthetic healthcare data generation. *IEEE Journal of Biomedical and Health Informatics.*

[18] Patel, R., Singh, A., & Kumar, S. (2024). Federated generative models for privacy-preserving medical data synthesis. *Artificial Intelligence in Medicine, 149*, 102749.