Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Blockchain-based NFT Framework for Ownership and Intellectual Property Preserving with Controlled Access to Medical Datasets

Hoda Naseri
Tarbiat Modares University
Seyed Mohammad Mirhosseini
Tarbiat Modares University
Ali A. Safaei
aa.safaei@modares.ac.ir

Tarbiat Modares University

Article

Keywords: Medical Data Sharing, Ownership preserving, Blockchain, NFT, Access Control, Intellectual property preserving, Leakage Detection

Posted Date: June 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9952836/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on October 3rd, 2026.
See the published version at https://doi.org/10.1038/s41598-026-67808-z.

A Blockchain-based NFT Framework for Ownership and Intellectual Property Preserving with Controlled Access to Medical Datasets

Hoda Naseri[1], Seyed Mohammad Mirhosseini[1], Ali A. Safaei[1,2*]

[1] Department of Data Science, Faculty of Interdisciplinary Science & Technology, Tarbiat Modares University, Tehran, Iran

[2] Department of Medical Informatics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran

* Corresponding author: Ali A. Safaei: aa.safaei@modares.ac.ir

Abstract

The rapid diffusion of artificial intelligence (AI) and advanced data analytics techniques across biomedical research, diagnostics and personalized medicine has established high-quality medical datasets as foundational resources. However, a significant impediment to progress is the reluctance of data owners to share these valuable resources because existing infrastructures provide no reliable guarantee of ownership or intellectual property preservation, in addition to the control over unauthorized usage. Consequently, vast quantities of recorded, high-potential data remain unused within private repositories. To overcome this barrier, a blockchain-based framework leveraging Non-Fungible Tokens (NFTs) is introduced in this paper to preserve ownership of medical datasets. The general workflow involves minting an NFT representing the dataset (stored encrypted off-chain), allowing users to request time-limited access via a smart contract. Access is granted using ephemeral decryption keys, enforcing fine-grained and revocable control. For forensic auditing, watermarking and perceptual hashing are integrated to enable the detection of post-access leakage without identifying the perpetrator directly. Technically, the framework integrates on-chain Merkle Tree-based integrity verification to ensure data fidelity upon retrieval. Practical applicability is demonstrated using a longitudinal multimodal neuroimaging dataset from OpenNeuro (DS007328). The proposed framework establishes a technically feasible and reproducible model for next-generation medical data governance, enabling verifiable ownership, controlled access, and reproducible auditing while keeping sensitive data off-chain.

Keywords: *Medical Data Sharing, Ownership preserving, Blockchain, NFT, Access Control, Intellectual property preserving, Leakage Detection*

1. Introduction

The rapid growth of medical imaging and neurophysiological datasets, including MRI, fMRI, EEG, iEEG, and transcranial magnetic stimulation (TMS) recordings, has significantly advanced biomedical research and data-driven healthcare. [14] [15] These datasets are fundamental to clinical studies, neuroscience

research, and the development of machine learning models for diagnosis, prognosis, and treatment planning. Despite their increasing value, preserving ownership, enforcing access rights, and ensuring accountability for medical data remain open challenges, particularly in collaborative and multi-institutional research environments. [16]

Most existing medical data repositories rely on centralized infrastructures governed by institutional policies. [17] While such systems enable controlled data distribution, they provide limited transparency and weak guarantees once access is granted. After a dataset is shared with a researcher, data owners typically lose visibility into how, when, and for what purpose the data are used. In cases of unauthorized redistribution or data leakage, establishing cryptographically verifiable ownership or tracing the origin of the breach is often infeasible. [18] These limitations undermine trust among data providers and act as a barrier to large-scale data sharing and reuse. [19]

Blockchain technology has emerged as a promising paradigm for improving transparency, immutability, and auditability in data management systems. [20] In parallel, Non-Fungible Tokens (NFTs) have attracted attention as a mechanism for representing unique digital assets and ownership claims on blockchain networks. [21] While NFTs have been widely adopted for digital art and intellectual property, their applicability to sensitive medical datasets remains largely unexplored. [22] Directly storing medical data on-chain is impossible due to privacy, scalability, and regulatory constraints, and existing NFT models alone do not address access control, data confidentiality, or post-access misuse. [6] Consequently, current medical data sharing platforms lack cryptographically verifiable ownership, fine-grained access enforcement, and forensic traceability after disclosure. [23]

A comprehensive solution for medical data ownership preservation must therefore go beyond simple tokenization. [6] Such a system must integrate off-chain encrypted storage, cryptographic integrity verification, fine-grained and revocable access control, immutable usage auditing, and post-disclosure accountability mechanisms. [31] Equally important, the system must remain practical for real-world stakeholders by addressing key management, usability, and recovery challenges.

A fundamental barrier to medical data sharing is the absence of a verifiable, non-repudiable ownership preservation mechanism that treats datasets as formal digital assets. [32] If medical datasets could be technically and legally recognized as protected digital assets, with enforceable ownership guarantees, data providers would no longer perceive sharing as a loss of control over intellectual or economic rights. [33]

A robust ownership-preserving framework must provide such properties:

- *Non-repudiation of ownership*
- *Immutability of ownership records*
- *Cryptographically verifiable provenance*
- *Traceability of usage and potential leakage*
- *Legally admissible digital evidence*

Such a mechanism enables ownership enforcement both legally, through intellectual property frameworks, and operationally, through verifiable transaction histories in cross-institutional environments. [34] [35]

To address these requirements, this paper proposes a blockchain-based NFT framework for secure medical data ownership, controlled access, and leakage traceability. In the proposed approach, each medical dataset is represented by an NFT whose metadata includes a Merkle Root derived from encrypted off-chain data stored in decentralized storage systems such as IPFS or Filecoin. Smart contracts govern access requests, approvals, and logging, while ephemeral, time-bound decryption keys enforce cryptographic access control.

To support post-access accountability, digital watermarking and perceptual fingerprinting techniques are employed to enable leakage detection and attribution. Furthermore, a custodial blockchain service with identity verification is incorporated to support secure key recovery and operational reliability. (Fig. 1)

Although prior studies have combined blockchain, NFTs, Merkle Trees, and decentralized storage for data provenance,[36] [37] the novelty of the proposed framework lies not in isolated technical components, but in their systematic integration into a unified governance architecture tailored for regulated medical data. [38] Specifically, this work introduces (i) a formal separation between full ownership transfer and licensed access delegation encoded directly at the smart contract level, (ii) a forensic reconstruction layer that links watermark attribution, Merkle membership proofs, and access-event logs into a traceable propagation chain, (iii) an on-chain evidence packaging mechanism that generates cryptographically verifiable bundles suitable for compliance review or legal enforcement, and (iv) a custodial key recovery architecture natively integrated with NFT-based ownership governance. Collectively, these elements extend the functional role of NFTs from passive provenance indicators to programmable governance primitives capable of integrating ownership representation, access lifecycle management, and post-distribution accountability within a unified cryptographic framework.

A natural question is whether similar functionality could be achieved using a centralized Public Key Infrastructure (PKI) combined with conventional audit logging. [39] While centralized systems can enforce access policies, they rely on institutional trust assumptions and allow retrospective modification of logs by privileged administrators. [40] In contrast, the proposed blockchain-based architecture provides cryptographically enforced ownership binding and tamper-resistant audit trails that are verifiable independently of any single institutional authority. The Merkle-root anchoring of encrypted datasets establishes immutable integrity commitments, while smart contracts encode governance rules in a transparent and deterministic manner. In cross-institutional and multi-jurisdictional research collaborations, where stakeholders may not share a common trusted intermediary, such decentralized verifiability reduces reliance on organizational trust and strengthens evidentiary robustness. Thus, the use of blockchain is not merely architectural preference but a deliberate design choice to achieve cryptographic non-repudiation, transparent governance, and independent auditability in sensitive medical data ecosystems. This study presents an architectural and methodological framework intended to formalize secure ownership and access governance for medical datasets, rather than a fully deployed clinical infrastructure.

The main contributions of this work are summarized as follows:

1. ***NFT-based ownership model for medical datasets*** *that provides immutable and verifiable proof of ownership and dataset integrity through Merkle Tree commitments and decentralized storage.*
2. ***Smart contract–driven access control framework*** *enabling permissioned and auditable data sharing without transferring ownership.*
3. ***Secure ephemeral key distribution mechanism*** *that enforces time-bound and single-use access to encrypted medical data.*
4. ***Integrated leakage detection and forensic evidence framework*** *based on watermarking, perceptual hashing, and blockchain-based audit logs.*
5. ***Custodial blockchain architecture*** *that addresses practical usability concerns through secure key management and recovery mechanisms.*

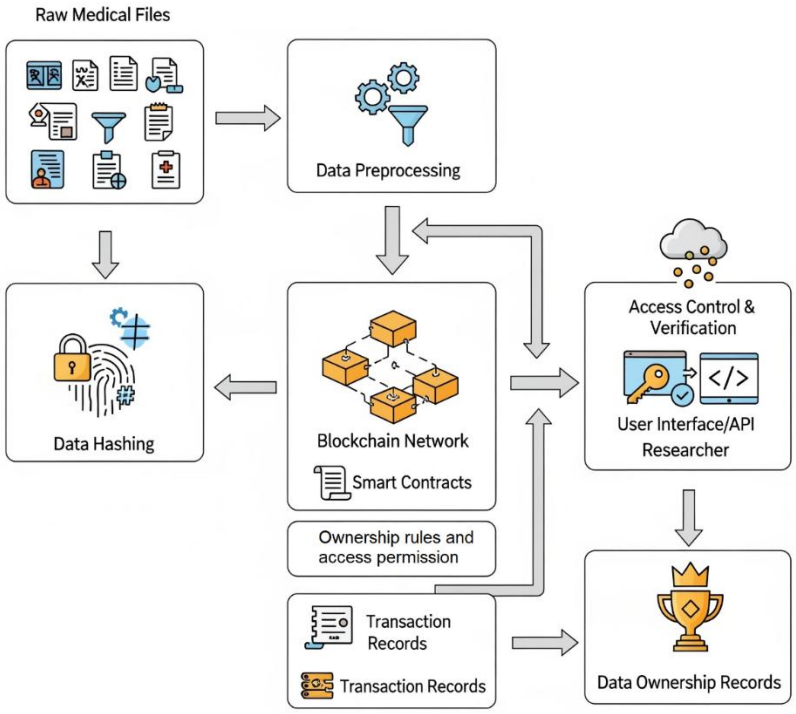


Fig. 1 the process of medical data ownership preservation using the proposed framework

The diagram (Fig. 1) illustrates a blockchain-based framework for secure and transparent management of medical data. Initially, raw medical files undergo data preprocessing to ensure consistency, accuracy, and quality. The processed data is then subjected to data hashing, generating unique cryptographic fingerprints that protect the integrity and confidentiality of patient information. These hashed data points are incorporated into a blockchain network, where smart contracts enforce ownership rules and access permissions autonomously. Every interaction and modification within the network is recorded as transaction records, ensuring traceability and accountability. Authorized users, such as researchers, access the data through a user interface or API, where **access** control and verification mechanisms authenticate and validate their permissions. Ultimately, all verified and authorized activities are compiled into data ownership records, establishing a transparent and tamper-proof system for managing medical data ownership and sharing.

The remainder of this paper is organized as follows. Section 2 reviews related work. Section 3 presents the proposed framework, including architecture, actor roles, dataset lifecycle, and end-to-end protocols. Section 4 provides security and threat analysis with formal verification of four core properties. Section 5 evaluates practical deployment and performance using the OpenNeuro dataset. Section 6 discusses limitations and trade-offs. Section 7 concludes and outlines future work.

2. Background and Related Work

Secure medical data sharing has received increasing attention with the rise of data-driven healthcare and the enforcement of strict privacy regulations such as HIPAA and GDPR [1]. A significant body of research has explored blockchain technology as a means to enhance transparency, immutability, and auditability in medical data management systems [2]. Early blockchain-based approaches primarily stored metadata, access logs, or consent records on-chain, while retaining raw medical data in centralized databases or cloud

storage [2][3]. In these systems, blockchain acts as a tamper-resistant ledger for access events and consent management, improving trust among stakeholders. [24]

More recent studies adopt hybrid architectures that combine blockchain with off-chain storage. In these designs, encrypted medical data are stored externally, while cryptographic hashes or pointers are recorded on-chain to ensure data integrity and provenance [4]. Smart contracts are often used to automate consent management and access control policies. [25] Although such approaches improve auditability, they generally lack mechanisms for post-access accountability. Once access is granted, tracing unauthorized redistribution or data leakage remains difficult. [26] Moreover, access permissions are frequently static or long-term, which violates the principle of least privilege and increases the risk of abuse. Key management is typically assumed to be secure at the user side, overlooking real-world challenges such as key loss, compromise, or recovery [5].

To enable fine-grained access control, cryptographic techniques such as attribute-based encryption (ABE), proxy re-encryption (PRE), and secure multi-party computation (MPC) have been applied to medical data sharing [6]. ABE allows data owners to define access policies based on user attributes, while PRE enables ciphertext transformation for authorized users without exposing plaintext. [27] Although these schemes provide strong confidentiality guarantees, they often introduce significant computational overhead and complex key management requirements. For large-scale datasets containing thousands of files, policy updates or re-encryption operations can become prohibitively expensive. [41] Furthermore, most cryptographic access control frameworks focus on preventing unauthorized access but do not address post-decryption misuse. [42] [43] Once plaintext data are accessed, built-in mechanisms for leakage tracing are generally absent.

Key Management Services (KMS) and MPC-based custodial solutions have been proposed to mitigate key-loss risks and improve operational reliability [7]. However, these solutions are often deployed as standalone components and are not tightly integrated with blockchain-based ownership representations or immutable audit trails.

Non-Fungible Tokens (NFTs) have been widely studied as a mechanism for representing ownership, provenance, and authenticity of digital assets on blockchain networks, particularly in the domains of digital art, collectibles, and intellectual property [8] [44]. Several works extend NFTs to scientific datasets by encoding metadata, ownership claims, and licensing terms [9]. Nevertheless, most NFT-based data models store minimal metadata on-chain and rely on off-chain trust for data integrity and access enforcement. In the context of medical data, naïve NFT implementations are insufficient due to privacy, scalability, and regulatory constraints [10]. Current NFT standards lack native support for fine-grained access control, time-limited permissions, and cryptographic evidence generation in cases of misuse. [45]

Merkle Trees are a well-established cryptographic primitive for integrity verification and membership proofs, widely used in blockchain protocols, distributed storage systems, and secure auditing frameworks [11] [46]. Several medical data management systems employ Merkle Trees to verify off-chain datasets. However, in most existing approaches, Merkle Roots are used solely for integrity checking and are not treated as ownership-binding or legally meaningful identifiers. As a result, dataset integrity, ownership proof, and access rights are not unified within a single on-chain artifact [12] [47].

In summary, existing solutions fail to simultaneously provide cryptographically verifiable ownership, fine-grained access control, post-access leakage attribution, and practical key recovery. Blockchain-based sharing improves auditability but lacks post-disclosure accountability; cryptographic access control ensures confidentiality but does not prevent misuse after decryption; and NFT-based ownership models offer provenance without addressing the unique requirements of sensitive medical data.

This paper addresses these gaps by proposing an integrated NFT-based framework that unifies Merkle Tree–based integrity verification, smart contract–driven access control, ephemeral key issuance, leakage detection, and custodial key recovery. Unlike prior work, the proposed system treats NFTs as comprehensive ownership and audit artifacts that bind dataset integrity, access history, and cryptographic evidence within a single immutable on-chain representation.

3. The Proposed Framework and Workflow

The proposed framework integrates blockchain-based NFTs with encrypted off-chain storage to establish three core capabilities: ownership preservation, controlled access, and post-access accountability.

Ownership Preservation: Each dataset is represented by a unique NFT whose metadata embeds a Merkle Root derived from encrypted files. This root functions as a cryptographic integrity commitment.

Controlled Access Governance: Access requests are processed via smart contracts. Approved researchers receive time-limited ephemeral decryption keys cryptographically bound to their identities and access scope.

Post-Access Accountability: Watermarking and perceptual fingerprinting enable forensic attribution of leaked content. All access and governance events are recorded on-chain.

Operational Recoverability: Custodial MPC-based key management and guardian approval workflows support secure key recovery.

The architecture integrates: *(i)* a public blockchain layer for ownership and auditability, *(ii)* decentralized storage (IPFS/Filecoin), *(iii)* AES-256 authenticated encryption, *(iv)* binary Merkle Tree integrity structures, *(v)* ERC-721/ERC-1155 compliant smart contracts, (vi) watermarking and fingerprinting modules, and *(vii)* custodial key recovery services. NFTs function as cryptographic ownership and governance artifacts.

The proposed NFT-based framework is designed to preserve dataset ownership, verify data integrity, enforce controlled access, and enable post-distribution accountability for medical datasets. Sensitive medical data are never stored on-chain; instead, encrypted datasets reside in decentralized storage networks, while NFTs and smart contracts provide immutable ownership representation, programmable access governance, and auditability. The architecture integrates:

- *Public blockchain layer (ownership + auditability)*
- *Decentralized storage*
- *AES-256 authenticated encryption*
- *Binary Merkle Tree integrity structure*
- *Smart contracts (ERC-721 / ERC-1155 compliant)*
- *Watermarking and fingerprinting modules*
- *Custodial key recovery services*

NFTs function as cryptographic ownership and governance artifacts. Using the OpenNeuro dense longitudinal single-subject MRI dataset as a reference example, the workflow demonstrates dataset encryption, Merkle Tree construction, NFT minting, ownership transfer, and access management without modifying the underlying data.

3.1. System Scope

The system (as a mechanism) is designed for high-value medical and biomedical datasets, including neuroimaging data (MRI, fMRI), electrophysiological signals (EEG, iEEG, TMS), clinical datasets, pathology data, OMICS data and etc. This framework can operate in two ways: either as a governance module within an Open Data Platform, or as a standalone tokenization tool that converts a dataset into an ownership-bound NFT artifact.

3.2. Architectural Layers and Functional Components

Figure 2 presents the component diagram of the proposed NFT-based medical data governance framework, comprising six interconnected layers that collectively address the requirements of secure medical data sharing. The architecture follows a vertical pipeline from data preparation at the top to monitoring and recovery at the bottom, with external actors interacting at specific layers.

Layer 1 – Data Curation and Preparation: Prior to registration, datasets undergo anonymization and de-identification to remove personally identifiable information from medical imaging and neurophysiological recordings, including MRI, fMRI, EEG, and iEEG data. To enable post-access accountability, each file is augmented with a robust digital watermark or perceptual fingerprint uniquely bound to the intended consumer. The prepared files are then encrypted using AES-256 in authenticated encryption mode (e.g., GCM) before leaving the data owner's secure environment.

Layer 2 – Decentralized Storage: Encrypted dataset files are uploaded to decentralized storage networks such as IPFS or Filecoin. Each file produces a content identifier (CID), which uniquely references the encrypted content. Cryptographic hashes of all encrypted files are computed using SHA-256 and aggregated into a Merkle Tree. The resulting Merkle Root acts as a compact 32-byte cryptographic commitment to the entire dataset and is later anchored on-chain. This Merkle Root ensures dataset integrity verification, tamper detection, and establishes a cryptographic link between off-chain data and on-chain ownership.

Layer 3 – NFT Tokenization and Ownership Representation: Each dataset is represented by a unique NFT implemented using ERC-721 or ERC-1155 standards. The NFT metadata includes the Merkle Root, CID list reference (or off-chain metadata pointer), dataset description, licensing terms, the owner's blockchain address, and cryptographic signatures. The NFT functions as an immutable and verifiable proof of dataset ownership and integrity, serving as the foundational governance primitive for all subsequent operations.

Metadata (json):

```
{
  "name": "OpenNeuro DS007328 MRI Dataset",
  "description": "85 hours resting-state fMRI, 458 runs, 243 sessions",
  "merkle_root": "0xabc123def456...",
  "cid_list_uri": "ipfs://QmListOfAllCIDs",
  "license": "CC BY-NC 4.0",
  "owner": "0xHospitalWallet",
  "signature": "0xSignatureOfOwner",
  "created_at": "2025-01-15T10:00:00Z"
}
```

Smart Contract Function (solidity):

```
function mintDatasetNFT(
    bytes32 merkleRoot,
    string memory metadataURI,
    address owner
```

```
) external returns (uint256 tokenId) {
    tokenId = _nextTokenId++;
    _safeMint(owner, tokenId);
    datasetMerkleRoots[tokenId] = merkleRoot;
    _setTokenURI(tokenId, metadataURI);
    emit DatasetMinted(tokenId, merkleRoot, owner, block.timestamp);
}
```

Layer 4 – Transaction and Ownership Transfer Model: The framework distinguishes between full ownership transfer and licensed access delegation, each governed by distinct smart contract operations. *Full ownership transfer* is facilitated via on-chain reassignment of NFTs, simultaneously registering transfer intent, settling terms, and transmitting the NFT to the new holder. Cryptographic keys are immediately rotated to nullify prior access permissions, leaving an unalterable on-chain log of the ownership transition. The dataset's integrity remains intact throughout, as Merkle Roots stay consistent. *Licensed access delegation* involves no NFT transfer but grants time-bound, scope-limited usage rights, with all events immutably recorded. *Authorized sub-licensing* supports controlled redistribution under programmable constraints (e.g., depth limits, royalty rules, notification requirements), where new access credentials are derived under smart contract supervision without direct forwarding of decryption keys.

Layer 5 – Access Control and Key Management: Ownership and access rights are explicitly separated: ownership is encoded in the NFT, while usage rights are granted through permissioned access records managed by smart contracts. Access requests, approvals, and revocations are recorded on-chain. Approved researchers receive ephemeral, time-limited decryption keys via a Key Management Service (KMS) or proxy re-encryption mechanism. These keys satisfy four criteria: bound to specific consumer identities, restricted to predefined datasets, time-limited, and revocable at any moment.

Layer 6 – Monitoring, Auditing, and Custodial Recovery: Post-access accountability is supported through continuous monitoring of public networks and repositories. Monitoring agents perform CID matching, perceptual hash (pHash) comparison, and watermark extraction to detect unauthorized dissemination. When suspected leakage is detected, a cryptographically verifiable evidence package is generated, linking the leaked artifact, the original dataset Merkle Root, the NFT ownership record, and the most recent authorized access event — enabling traceable attribution for compliance, auditing, or legal proceedings. Detection also triggers access suspension, license revocation, and on-chain audit recordings. To address usability and operational risks, cryptographic keys are protected using MPC-based custody or secure KMS solutions. Key recovery requires off-chain identity verification (e.g., KYC), multi-party guardian approval (e.g., 3-of-5 threshold), and on-chain recording of recovery events, preserving transparency and accountability throughout the recovery lifecycle.

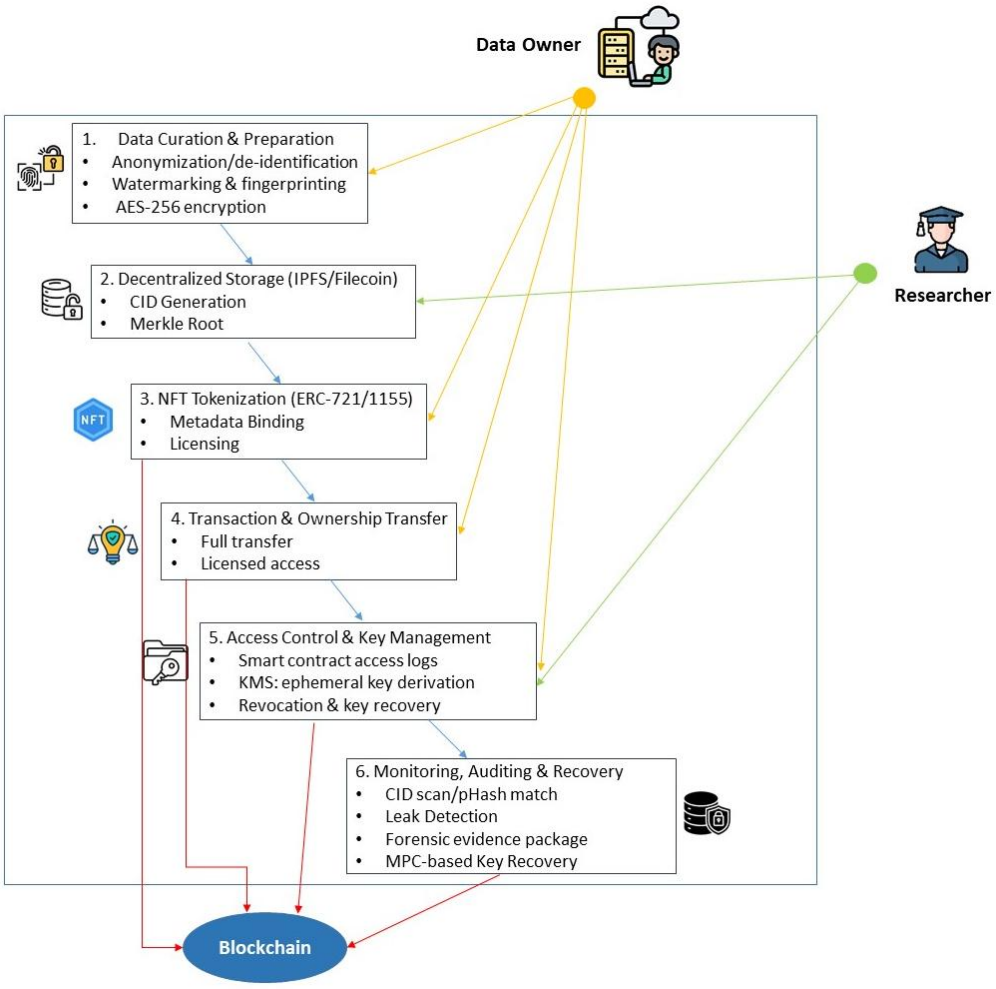


Fig. 2 Component diagram of the six-layer architecture

As shown in Figure 2, the six layers are vertically interconnected, with encrypted data and Merkle commitments flowing downward from preparation through storage and tokenization to access control and monitoring. External actors — the Data Owner, Researcher, and Blockchain Network — interact with specific layers as indicated by the arrows, ensuring that sensitive medical data never reside on-chain while maintaining cryptographic ownership, fine-grained access control, and post-distribution forensic accountability.

3.3. Actors and Roles

Figure 3 illustrates the use case diagram of the proposed framework, depicting seven actors and their interactions. Two human actors are shown: the Data Owner (mints NFTs, uploads encrypted data, approves/revokes access, transfers ownership) and the Researcher (requests access, downloads via CID, receives ephemeral keys). Five system components complete the architecture: the Ownership Verification Platform (OVP) (smart contract governance for NFT lifecycle and access logs); the Blockchain Network (BC) (immutable ledger for ownership records, transfers, and events); the Decentralized Storage (DS) (IPFS/Filecoin for encrypted off-chain storage and CID resolution); the Key Management Service (KMS)

(ephemeral key derivation and MPC-based recovery); and the Monitoring Agent (MA) (CID/pHash leak detection, watermark extraction, forensic evidence generation). Solid arrows indicate primary interactions (e.g., NFT minting, access requests, transaction recording), while dashed arrows represent supporting flows (e.g., ephemeral key delivery from KMS to Researcher). This actor ensemble ensures end-to-end governance without storing sensitive data on-chain.

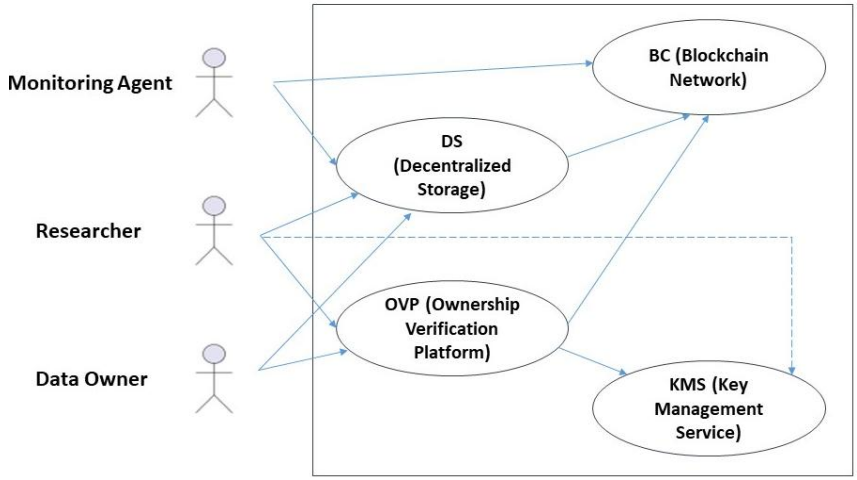


Fig. 3 Use case diagram of actors in the NFT-based medical data governance framework

3.4. Design Rationale and Security Assumptions

The design ensures that sensitive medical data are never stored on-chain and that cryptographic mechanisms enforce confidentiality and access control. By combining NFTs, Merkle-based integrity verification, ephemeral key access, and continuous monitoring, the system enables verifiable post-access accountability and leakage traceability. The architecture balances decentralization with operational usability by incorporating custodial safeguards.

Security assumptions include partial trust in platform operators for off-chain services, the immutability of blockchain records, and the cryptographic security of underlying primitives.

3.5 Dataset Lifecycle: Stage-by-Stage Description

The journey which a dataset will experience through the proposed framework is described in this section.

Fig. 4 illustrates the complete seven-stage dataset lifecycle. Stage 1 (Creation): The Data Owner anonymizes (PHI removal, $k \geq 5$), watermarks (DWT-SVD, 128-bit), encrypts (AES-256-GCM with key K_d), uploads to IPFS (CID generation), constructs a Merkle Tree (SHA-256 → root MR), mints an NFT (mintDatasetNFT), and shards K_d (3-of-5 Shamir). Stage 2 (Request): The Researcher discovers the dataset, verifies MR on-chain, and submits a signed request $\sigma = \text{sign}(sk_R, \text{tokenId} \parallel \text{purpose} \parallel \text{duration} \parallel \text{nonce})$; the contract logs AccessRequested. Stage 3 (Key Issuance): The Data Owner calls approveAccess(), the KMS derives $K_e = \text{SHA-256}(K_d \parallel R_{pub} \parallel \text{nonce} \parallel \text{expiry} \parallel \text{tokenId})$ (identity-bound, time-limited, revocable), encrypts it with the researcher's public key, and delivers via TLS 1.3. Stage 4 (Usage): The Researcher fetches CIDs, downloads via ipfs get <CID>, verifies with verify(MR, SHA256(ciphertext), π), decrypts

(AES-256-GCM), and analyzes. Disclosure: Offline copying cannot be prevented; the framework provides accountable traceability via watermarks and on-chain logs. Stage 5 (Transfers): Three scenarios exist – full transfer (safeTransferFrom(), rotates K_d'), licensed access (requestAccess()/approveAccess(), no NFT transfer, ephemeral K_e), and sub-licensing (createDerivedAccess(), derived K_e'). Keys are NEVER forwarded; each user receives uniquely watermarked copies. Stage 6 (Monitoring): Access auto-expires (block.timestamp > expiry); the Monitoring Agent scans using CID matching (100% precision), perceptual hashing (95% under JPEG Q70), and watermark extraction. Stage 7 (Response): Upon leak detection, the agent extracts the watermark, queries blockchain logs, generates an evidence package {evidence_id, dataset {token_id, merkle_root}, leaked_file {cid, watermark_signature}, attribution {consumer_wallet, license_id}, blockchain_proofs}, and the contract auto-calls revokeAccess(). The package supports license revocation, notification, regulatory reporting, and legal proceedings.

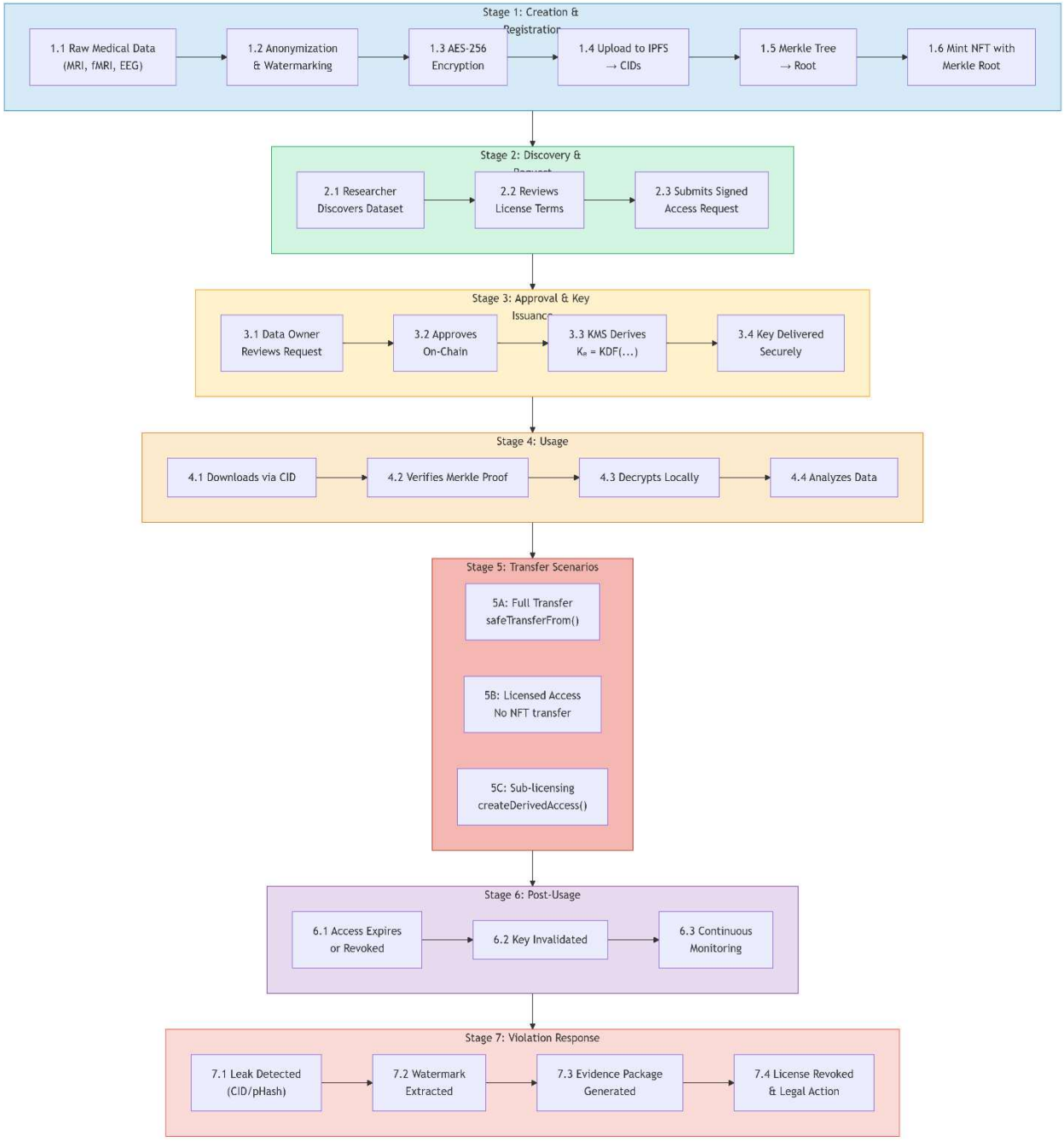


Fig. 4 the complete dataset lifecycle

3.6 End-to-End Protocols

This section formalizes the complete lifecycle of dataset registration, access control, ownership transfer, redistribution, and forensic accountability within the proposed NFT-based governance framework (Fig. 6).

Step 1: Dataset Registration and NFT Minting

Dataset registration begins with anonymization and watermarking of the dataset by the Data Owner. Each file is encrypted using a dataset-specific symmetric key K_d. The encrypted files are uploaded to decentralized storage, generating a set of content identifiers:

$$\text{CID} = \{Cid_1, Cid_2, \ldots, Cid_n\}$$

Cryptographic hashes of the encrypted files are computed and aggregated into a Merkle Tree:

$$\text{SHA256}\ (file_i) = h_i, \qquad \text{MR} = \text{MerkleRoot}\ (h_1, h_2, \ldots, h_n)$$

The Data Owner then assembles the NFT metadata containing the Merkle Root, dataset descriptors, licensing information, ownership address, and digital signatures. By invoking `mintDatasetNFT(metadata)` on the smart contract, the NFT is minted and recorded on-chain. The emitted `DatasetMinted` event establishes the NFT as the canonical and immutable proof of dataset ownership.

mintDatasetNFT(MR, metadataURI, owner) → tokenId

Event: DatasetMinted(tokenId, MR, owner, timestamp)

Step 2: Access Request and Authorization

Researcher submits a signed access request specifying the dataset identifier, intended purpose, and access duration. The request is logged on-chain via `requestAccess(tokenId)`:

$$\text{Req} = sign_R(tokenId, purpose, duration)$$

The Data Owner evaluates the request off-chain (e.g., ethical approval) and records the decision on-chain using either `approveAccess(requestId)` or `denyAccess(requestId)`. The blockchain immutably records the decision, ensuring transparency and auditability.

requestAccess(tokenId, purpose, duration, Req) → requestId

approveAccess(requestId) OR denyAccess(requestId)

Event: AccessApproved(requestId, licenseId, researcher, expiry)

Step 3: Ephemeral Key Issuance and Secure Access

Upon approval, the KMS derives a one-time decryption key:

$$K_e = KDF(K_d, R_{pub}, nonce, expiry)$$

Where: K_d= dataset master key, R_{pub} = researcher public key, nonce = randomness, expiry = validity window

The key is cryptographically bound to researcher identity, dataset NFT and validity period. Delivery occurs via MPC-based secure channels or decentralized key access protocols.

The Researcher downloads encrypted files from DS and decrypts them locally:

$$\text{File} = AES_{K_e}^{-1}(\text{ciphertext})$$

A `DataDecrypted` event is emitted, linking usage to the researcher and NFT.

$$\text{deliver}(K_e, \text{researcher}) \text{ via secure channel}$$
$$\text{Event: KeyIssued(licenseId, researcher, expiry)}$$

Step 4: Ownership Transfer and Dataset Sharing

The framework distinguishes three transaction categories:

A. Full Ownership Transfer

NFT reassignment using:

```
initiateTransfer(datasetId, newOwner, termsHash)
safeTransferFrom(A, B, tokenId)
```

Following transfer:

- Master key rotation is executed
- Previous access credentials are invalidated
- Merkle Root remains unchanged
- `OwnershipTransferred` event is emitted

This preserves dataset integrity commitments while resetting access governance.

B. Licensed Access (Non-Ownership Sharing)

Access licenses encode:

```
{
  "dataset_id": "ds007328",
  "consumer": "0xABC...",
  "scope": "sub-001/ses-2024110602/func",
  "purpose": "Non-commercial research",
  "expiry": "2027-01-01",
  "revocable": true
}
```

Access flow:

1. `requestAccess(tokenId, scope, purpose)`
2. Owner review
3. `approveAccess(requestId)`
4. Ephemeral key derivation
5. Time-limited access window
6. Access logging

Ownership remains unchanged.

C. Authorized Secondary Redistribution (Sub-licensing)

If policy permits:

```
allowSublicense = true
maxDepth = d
%royalty = r
```

Delegation is performed via: `createDerivedAccess(newConsumer)`

Key forwarding is prohibited. Instead, derivative keys are generated under smart contract supervision. Royalty routing and owner notification occur automatically.

Step 5: Data Propagation Chain and Leakage Detection

Despite cryptographic controls, post-decryption copying cannot be entirely prevented. The system therefore enables reconstruction of propagation chains. (Fig. 5)



Fig. 5 Example of propagation chain

To reconstruct misuse:

Layer 1: Per-user watermarking

Each authorized user receives uniquely fingerprinted content.

Layer 2: On-Chain Footprint Traceability

The Footprint mechanism cryptographically binds four evidence categories into an immutable chain of custody: NFT transaction history (ownership transfers and license delegations), wallet addresses (actor identifiers), timestamped smart contract interactions (access requests, approvals, revocations), and dataset fingerprint metadata (Merkle roots, CID lists, watermarks, perceptual hashes). This enables any auditor to reconstruct the complete provenance and access propagation chain of a medical dataset without trusting centralized logs.

Layer 3: CID-based traceability

If leaked files appear on IPFS or public storage: CID match → hash match → Merkle proof verification

Monitoring agents continuously scan public networks for potential dataset leakage. Detected files are verified using CID matching, perceptual hashing, and watermark extraction. When a match is confirmed, a Merkle Proof is constructed to validate dataset membership:

$$\text{True} = \text{verify}(MR, h_i, \pi)$$

$$\text{Verify}(MR, \text{SHA256}(\text{leaked_file}), \pi) = \text{true} \Rightarrow \text{leaked_file} \in D$$

$$\text{extract_watermark}(\text{leaked_file}) \rightarrow \text{consumer_wallet}$$

$$\text{Evidence} = \{MR, \text{leaked_hash}, \text{watermark_sig}, \text{consumer}, \text{access_logs}, \text{blockchain_proofs}\}$$

Layer 4: Access event reconstruction

Attribution is derived from last access timestamp, scope information, embedded fingerprint signature, and on-chain logs.

Step 6: Forensic Evidence Package

When leakage is confirmed, a cryptographically verifiable evidence bundle is generated:

```
{
  "original_merkle_root": "...",
  "leaked_file_hash": "...",
  "watermark_signature": "...",
  "consumer_wallet": "...",
  "access_log_proof": "...",
  "timestamp_chain": "..."
}
```

This package supports license revocation, institutional notification, regulatory reporting and legal proceedings. The objective is deterrence and accountability rather than absolute technical prevention.

Step 7: Custodial Key Recovery

In case of key loss, users initiate a recovery request through `submitRecoveryRequest()`. After off-chain identity verification:

- Guardians approve via on-chain voting
- New key material is derived
- Previous keys are revoked
- Recovery event is permanently logged

$$\text{submitRecoveryRequest}(\text{user}) \rightarrow \text{recoveryId}$$

$$\text{approveRecovery}(\text{recoveryId}) \text{ by } \geq 3 \text{ of } 5 \text{ guardians}$$

$$\text{deriveNewKeyMaterial}()$$

$$\text{Event: RecoveryCompleted}(\text{user}, \text{timestamp})$$

This balances usability and cryptographic safety.

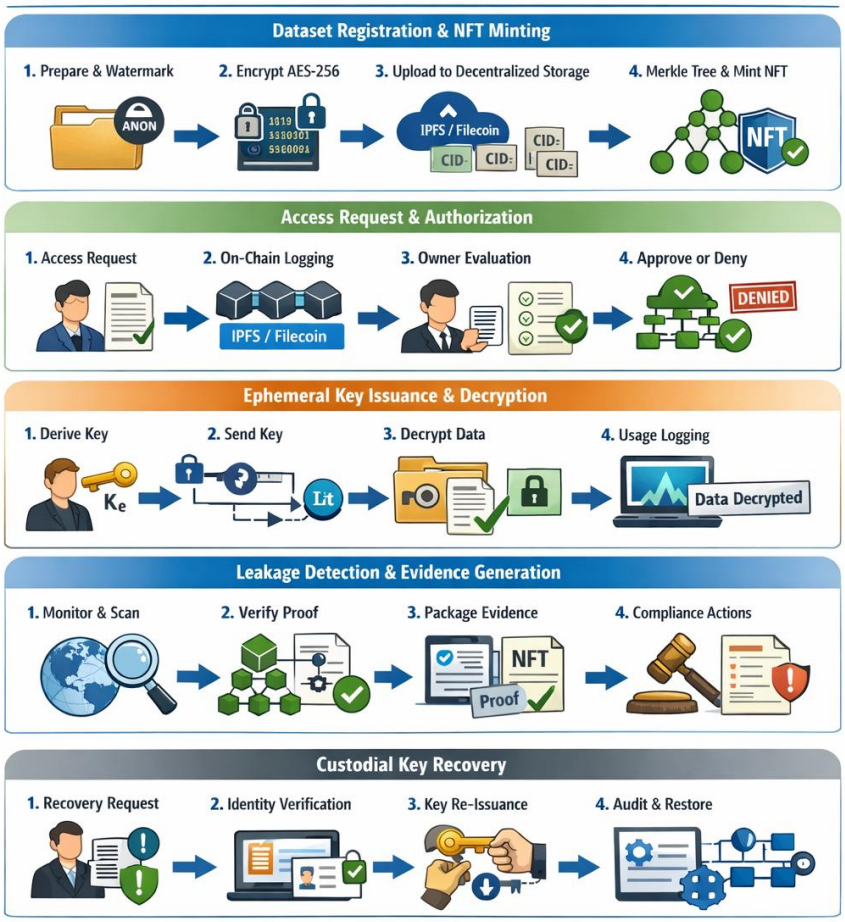


Fig. 6 End-to-End Protocols

The workflow (fig. 6) illustrates dataset encryption and Merkle commitment, NFT minting, on-chain access governance, ephemeral key derivation, ownership transfer, sub-licensing control, and post-distribution forensic accountability.

4. Security and Threat Analysis

This section analyzes the security properties of the proposed NFT-based medical data ownership framework. We evaluate the system against realistic adversarial models, identify attack surfaces, and describe corresponding mitigation mechanisms. The same access-control and ephemeral key mechanisms described here are directly applicable to the referenced neuroimaging dataset.

4.1. Security Objectives

The system is designed to satisfy the following security objectives:

- ***Confidentiality***: *Prevent unauthorized access to sensitive medical datasets.*
- ***Integrity***: *Ensure that datasets remain unaltered after registration.*
- ***Authenticity***: *Provide cryptographic proof of dataset ownership and origin.*
- ***Access accountability***: *Ensure that all dataset usage is attributable and auditable.*
- ***Non-repudiation***: *Prevent denial of ownership, access grants, or data usage.*

- ***Recoverability****: Enable secure recovery from key loss without compromising confidentiality.*
- ***Regulatory compliance****: Support legal and ethical requirements (e.g., HIPAA, GDPR).*

4.2. Threat Model

The proposed framework considers six distinct adversarial entities with varying capabilities and objectives. (Table. 1) External adversary (A_1) operates outside authorized channels, attempting unauthorized data access, impersonation of legitimate users, or traffic observation between system components. Malicious researcher (A_2) represents an insider threat who, despite holding legitimate access credentials, intentionally redistributes decrypted data, circumvents usage restrictions, or misuses the accessed medical information beyond declared purposes. Compromised platform insider (A_3) includes system administrators, KMS operators, or custodial guardians with privileged access who may attempt unauthorized key extraction, metadata manipulation, or collusion to bypass access controls. Blockchain attacker (A_4) targets the underlying distributed ledger through smart contract exploitation (e.g., reentrancy, integer overflow, or access control bypass) or consensus-layer attacks (e.g., 51% attack on proof-of-work or long-range attacks on proof-of-stake). Storage adversary (A_5) interacts with the off-chain decentralized storage layer, potentially observing encrypted data (passive) or tampering with ciphertexts, CIDs, or Merkle tree structures (active). Finally, key loss scenario (A_6) addresses non-adversarial but operationally critical failures where legitimate users lose private keys, wallet credentials, or recovery shards, potentially resulting in permanent data inaccessibility. The framework assumes standard cryptographic primitives remain secure and that the underlying blockchain consensus is not compromised at scale, while all adversarial actions are mitigated through mechanisms described in Sections 3.4 and 3.5.

Table. 1 Attack Surface and Mitigation Analysis

Attack Surface	Description	Mitigation Mechanisms
On-chain smart contracts	Ownership transfer, access logic	Audited libraries, minimal deterministic logic, role-based access control, reentrancy guards
NFT metadata	Dataset integrity and provenance	Immutable metadata, on-chain Merkle Root anchoring, cryptographic signatures
Off-chain storage	Encrypted data in IPFS/Filecoin	AES-256 encryption, CID safety, Merkle proofs, watermarking
Key management	Key generation, storage, recovery	MPC/KMS custody, ephemeral keys, guardian-based recovery
Access workflows	Request, approval, and key issuance	Signed requests, nonce/expiry enforcement, on-chain logging
Post-access usage	Unauthorized redistribution	Unique watermarks, monitoring agents, Merkle proof evidence, automated revocation

4.3. Formal specification and verification

This section presents a formal treatment of the security properties guaranteed by the proposed NFT-based medical data governance framework. We adopt a game-based security model and provide theorem-proving style derivations for four core properties: data confidentiality, integrity verification, ownership non-repudiation, and access accountability.

4.3.1. Data confidentiality

No adversary who is not explicitly authorized by an active, non-revoked access license shall be able to decrypt any encrypted medical dataset file with probability non-negligibly greater than random guessing. This property holds even if the adversary possesses arbitrary off-chain ciphertexts, observes on-chain metadata, or compromises up to t-1 of n MPC guardians (where threshold $t > n/2$).

$$\forall A \notin \text{authorized users} \rightarrow \Pr(A \text{ decrypts data}) \approx 0$$

$$\forall A, \forall C \leftarrow \text{Enc}(K_d, D), \forall t \in \text{Time}: \neg\text{Authorized}(A, t) \rightarrow \Pr[A(C, \text{view}) = D] \leq \tfrac{1}{2} + \text{negl}(\lambda)$$

Where $\text{Enc}(K_d, D) \rightarrow C$ is AES-256-GCM encryption, $K_e = KDF(K_d, R_{pub}, nonce, expiry)$ is the ephemeral key derivation function, Authorized(t) indicates a valid unrevoked license, and negl(λ) is a negligible function in security parameter λ.

Theorem 1 *(Data Confidentiality):* Under IND-CCA2 security of AES-256-GCM and KDF modeled as a random oracle, confidentiality holds against adversaries A_1–A_5 who lack a valid ephemeral key K_e.

> *Proof Sketch:* Any adversary without K_e cannot distinguish ciphertext from random (reduction to AES-256-GCM). For A_3, threshold secret sharing ensures < t guardians have zero information about K_d. For A_2 post-revocation, key destruction reverts capabilities to A_1.

□

4.3.2. Integrity verification

A file verifies against the on-chain Merkle root if and only if it is an unmodified member of the original dataset.

$$\forall f, \forall MR: \text{Verify}(MR, \text{SHA256}(\text{Enc}(K_d, f)), \pi) = \text{true} \Leftrightarrow f \in D$$

Where $D = \{f_1,...,f_n\}$ is the dataset, $h_i = \text{SHA256}(\text{Enc}(K_d, f_i))$, $MR = \text{MerkleRoot}(h_1, \ldots, h_n)$, and π_i is the Merkle proof.

Theorem 2 *(Integrity Verification):* Under SHA-256 collision resistance, Merkle root anchoring provides completeness (authentic files verify) and soundness (tampered files fail verification).

> *Proof Sketch:* Completeness follows from Merkle tree construction. Soundness reduces to collision resistance: a false positive would require a leaf or internal node hash collision, which is negligible.

□

4.3.3. Ownership non-repudiation

NFT owners cannot deny minting or transferring ownership, as each event is immutably recorded and cryptographically signed.

$$\forall TX: \text{safeTransferFrom}(A, B, \text{tokenId}) \text{ on-chain} \wedge \text{Verify}(pk_A, TX) = \text{true} \Rightarrow A \text{ cannot deny transfer} \wedge B \text{ cannot deny receipt}$$

Theorem 3 *(Ownership Non-Repudiation):* Under EUF-CMA security of ECDSA and blockchain immutability (honest-majority consensus), NFT ownership provides non-repudiation.

> *Proof Sketch:* Transactions are permanently recorded under honest-majority consensus. Signature unforgeability ensures only the legitimate key holder could

have signed. Cryptographic evidence prevents repudiation (modulo key compromise, addressed in Section 4.4).

□

4.3.4. Access accountability

Every access event (request, approval, decryption, revocation) is recorded in an immutable on-chain log cryptographically bound to actor, dataset, and timestamp.

$$\forall \text{ action} \in \{\text{request, approve, revoke, decrypt}\}:$$

$$\text{action_executed}(A, \text{tokenId}, t) \Leftrightarrow \exists \text{ event } e \in \text{Log with } e.\text{actor} = A \wedge e.\text{tokenId} = \text{tokenId}$$

Theorem 4 *(Access Accountability):* If every critical smart contract function emits a unique event and blockchain logs are immutable, access accountability is satisfied.

Proof Sketch: All state-changing functions include emit statements. Ethereum logs are permanently stored in the receipt trie under honest-majority consensus. Events are cryptographically attributable via transaction signatures.

□

4.4. Residual Risks and Limitations

Despite cryptographic enforcement, certain residual risks remain. Client-side capture, such as screenshots or recordings of decrypted data, cannot be fully prevented. Legal enforceability of the framework's protections depends on jurisdictional recognition of NFTs as legally binding ownership instruments. Metadata access patterns, including the timing and frequency of on-chain interactions, may leak limited side-channel information about research activities. Finally, long-term cryptographic resilience requires future-proofing against advances in cryptanalysis, particularly the potential emergence of quantum computing capabilities that could undermine current hash and signature schemes. (Table. 2)

Table. 2 Security Comparison with Existing Approaches

System	On-chain Ownership	Access Control	Leak Attribution	Key Recovery	Scalability	Performance Overhead
Centralized repository	✗	Policy-based	✗	Admin-only	Medium	Low
NFT registry only	✓	✗	✗	✗	High	Low
Proposed system	✓	✓	✓	✓	High	Moderate

The proposed architecture uniquely combines cryptographic access enforcement, post-access accountability, and secure key recovery, providing stronger end-to-end protection than prior approaches.

5. Practical Deployment and Performance Analysis

5.1 Dataset

To evaluate real-world feasibility, we conducted an end-to-end performance assessment using the publicly available neuroimaging dataset OpenNeuro **DS007328** (3,151 files) [13]. The objective was to measure computational overhead for integrity commitment, key derivation, encryption/decryption, scalability, and concurrent access under realistic research workloads. This dataset contains:

- *85 hours of resting-state fMRI across 458 runs (243 sessions)*
- *195 T1-weighted structural scans and 54 DTI sessions*
- *Detailed behavioral, physiological, and pharmacological metadata*

The dataset is fully anonymized and released under a CC0 license, making it suitable for illustrative NFT-based ownership and access workflows.

5.2 Experimental Evaluation

- **Full-Dataset Integrity Commitment**

All 3,151 files were hashed using SHA-256 and aggregated into a Merkle Tree to generate a single cryptographic commitment (Merkle Root) representing dataset integrity. (Table. 3)

Table. 3 Results of Integrity Commitment (Full Dataset)

Metric	Value
Files processed	3,151
Total hashing time	0.2194 s
Merkle Tree construction	0.0040 s
Proof generation	0.0093 ms
Proof verification	0.0196 ms
Verification result	True

Hash computation for the entire dataset was completed in under 0.22 seconds, while Merkle Tree construction required only 4 ms, and proof generation and verification were effectively instantaneous (<0.02 ms). The resulting Merkle Root, a fixed-size 32-byte value, provides a compact cryptographic commitment to the dataset's integrity, making it highly suitable for inclusion in on-chain NFT metadata and enabling efficient, verifiable ownership and access control.

- **Access Workflow Performance (Encryption & Key Derivation)**

To simulate controlled access, a benchmark was performed on 20 representative files including SHA-256 hashing, Merkle Tree construction, Ephemeral key derivation and AES-256 decryption.

Table. 4 Results of Access Simulation

Metric	Value
Files	20
Hashing time	0.0027 s
Merkle root	f5ca029f1b429256a40812b288e91ed29fa97910fe1070a308b775fbfcbb4ef8
Merkle build time	0.0001 s
Ephemeral key derivation	0.01 ms
AES-256 decryption (per file)	0.58 ms

Ephemeral key generation is effectively instantaneous, and AES-256 decryption requires less than 1 ms per file, allowing secure, time-bound access to be granted with near real-time responsiveness. (Table. 4) These

results confirm that the cryptographic access enforcement mechanisms of the framework introduce negligible latency, ensuring that end users experience seamless and efficient access to protected medical datasets.

- **Scalability Evaluation**

Subsets of increasing size were evaluated to empirically validate complexity behavior.

Table. 5 Results of Scalability Evaluation

Files	Hash Time (s)	Merkle Build (s)	Peak RAM (MB)
100	0.0171	0.0010	0.02
500	0.0435	0.0055	0.06
1000	0.0872	0.0126	0.12
2000	0.2278	0.0296	0.25
3151	0.3650	0.0253	0.39

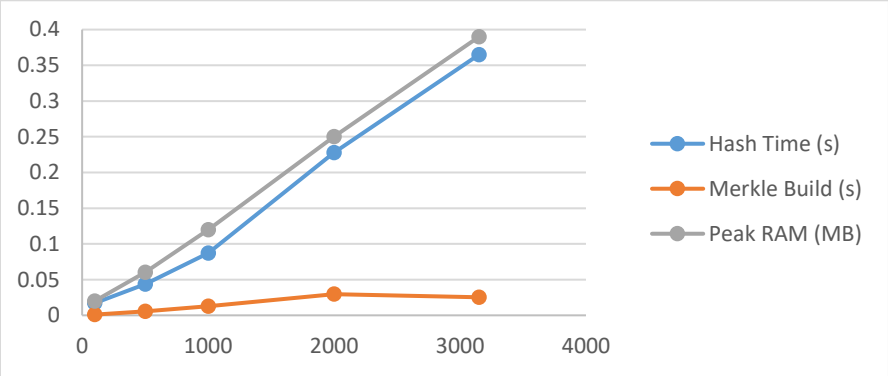


Fig. 7 Scatter Plot of Scalability Metrics

Hash computation demonstrates near-linear scaling with dataset size, while Merkle Tree construction remains computationally lightweight throughout. Memory usage stays below 0.5 MB even when processing all 3,151 files, and the total processing time for the full dataset remains under 0.4 seconds, confirming that the framework is highly efficient and suitable for large-scale medical data deployments. (Table. 5, fig. 7)

- **Empirical Complexity Validation**

Linear regression analysis produced:

- Hashing $R^2 = 0.992$
- Merkle construction $R^2 = 0.799$

Hashing confirms near-perfect *O(n)* behavior. Merkle construction shows minor timing variance at millisecond granularity due to measurement noise, but overall growth remains consistent with linear complexity.

- **Concurrent Access Simulation**

To simulate institutional workloads, ephemeral key derivation was executed concurrently.

Table. 6 Concurrent Access Simulation

Concurrent Users	Total Time (s)
10	0.0293
50	0.0277
100	0.0286
200	0.0329

Latency remained stable under 200 simultaneous requests, demonstrating strong parallel scalability and suitability for research environments with multiple active users. (Table. 6)

- **Forensic Leakage Reconstruction**

A multi-hop redistribution scenario was simulated by embedding unique watermark identifiers per user. When a redistributed copy was analyzed, the system successfully identified the responsible user:

Leak detected from User ID: 3

This confirms that the forensic accountability layer can reconstruct propagation chains and attribute misuse to specific access events.

- **Deployment Implications**

The empirical results indicate several key deployment characteristics. Dataset integrity commitment is near-instantaneous, while access authorization and decryption operate in real time. Memory and central processing unit overhead are minimal, and concurrent workloads do not degrade performance. Forensic traceability is operationally viable. Overall, the architecture is computationally lightweight and suitable for deployment in hospital research infrastructures, institutional data repositories, and regulated medical data governance environments. The constant-size Merkle Root commitment enables efficient NFT-based ownership binding, while off-chain encrypted storage preserves scalability. These properties collectively support practical, large-scale implementation of NFT-governed medical dataset management without prohibitive resource costs.

5.3 Deployment Considerations

The system can be deployed on Ethereum-compatible blockchain networks (e.g., Polygon or Avalanche), which offer scalability, low transaction fees, and fast confirmation suitable for NFT minting and access logging. Sensitive medical data are never stored on-chain; encrypted datasets reside in decentralized storage networks.

Key management is implemented using MPC-based custodial services or protocols such as Lit Protocol, enabling ephemeral key issuance, time-bound access, and secure recovery. Guardian-based recovery workflows ensure usability while preserving auditability. Users interact with the platform through lightweight web clients that support dataset registration, access requests, and secure local decryption.

5.4 Stakeholder Considerations

- **Data Owner Considerations**

Although the framework strengthens ownership guarantees, several operational challenges remain for data owners. Post-decryption control limitations mean that once plaintext data are legitimately accessed,

complete technical prevention of copying is infeasible. Governance overhead may increase administrative burden due to manual review of access requests. Jurisdictional variability affects legal recognition of NFT-based ownership claims, which differs across regulatory environments. Finally, master key custody risk remains: compromise of dataset master keys could undermine confidentiality without robust multi-party computation or custodial safeguards.

- **Dataset Consumer Considerations**

From the researcher's perspective, programmable governance introduces trade-offs. Metadata exposure risk means that on-chain access logs may reveal limited contextual information regarding research activity. Revocation risk exists because access privileges may be revoked under predefined policy conditions. Perceived monitoring may raise concerns about proportional oversight due to continuous accountability mechanisms. Compliance complexity may increase procedural overhead due to institutional review and licensing obligations. These trade-offs reflect an inherent tension between transparency, auditability, and privacy in blockchain-governed systems.

5.5 Economic and Operational Considerations

Deployment introduces additional economic and systemic factors, including transaction (gas) costs associated with NFT minting and access logging, migration and interoperability considerations across Layer-2 networks, royalty enforcement in secondary or delegated sharing, escrow dispute resolution in ownership transfers, and ethical implications of secondary data markets. The most important systemic risks in the use-case scenarios can be classified as follows: metadata leakage, where access timing and ownership transfers may reveal research trends; cross-jurisdictional regulatory conflicts, arising from divergence between frameworks such as GDPR and HIPAA; long-term decentralized storage persistence, including sustainability and pinning costs; and smart contract vulnerabilities, necessitating formal verification and external audits.

5.6 Theoretical Performance Analysis

The proposed framework exhibits predictable performance characteristics across its core operations. NFT minting latency is bounded by blockchain confirmation time; on Layer-2 networks, dataset registration typically completes within seconds. Access request and approval logic introduces negligible computational overhead, while on-chain logging ensures transparency with minimal latency. For a dataset of n files, Merkle Tree construction requires O(n) hash operations, while membership proof verification requires O(log n) time. Ephemeral key issuance, including one-time key derivation and secure delivery, can be completed within milliseconds per request. Data decryption using AES-256 incurs minimal overhead and scales linearly with dataset size.

5.7 Cost and Scalability Analysis

Only compact metadata, including Merkle Roots, ownership records, and access events, are stored on-chain, minimizing transaction costs. Batch processing and rollup mechanisms can further reduce these costs. Off-chain encrypted storage ensures that scalability is limited primarily by storage network throughput, while integrity verification scales logarithmically. Monitoring and leakage detection are performed off-chain and can be parallelized without affecting data access workflows. Although a full prototype implementation is beyond the scope of this study, theoretical analysis indicates that the proposed framework is practical, scalable, and suitable for real-world medical research environments. The hybrid on-chain and off-chain

design provides strong security guarantees while keeping computational and financial overhead manageable.

6. Discussion and Limitations

The increasing scale, heterogeneity, and regulatory sensitivity of medical datasets have exposed structural limitations in conventional data governance frameworks. [28] Traditional repositories provide secure storage and access control but rarely offer cryptographically verifiable ownership transfer, programmable enforcement of usage policies, or post-distribution accountability. [29] Conversely, most NFT-based systems focus on digital asset provenance without addressing the operational, ethical, and regulatory constraints inherent to healthcare data management. [30]

This study proposes an architectural bridge between these domains by treating NFTs not as speculative digital artifacts but as cryptographic governance primitives. In the proposed framework, each NFT is bound to encrypted medical datasets through Merkle-root commitments anchored on-chain. This hybrid design separates governance logic (on-chain) from sensitive data storage (off-chain), thereby preserving confidentiality while maintaining verifiable integrity and ownership traceability.

A key conceptual contribution is the formal distinction between ownership transfer and licensed access delegation. In regulated medical contexts, transferring legal custodianship of a dataset differs fundamentally from granting time-bound research access. Encoding this distinction at the smart contract level reduces governance ambiguity and enables auditable lifecycle management.

The integration of ephemeral key derivation and revocable, scope-limited access licenses further advances the state of medical data governance. Instead of static download permissions, access becomes programmable, identity-bound, and time-constrained. Although no technical system can fully prevent misuse after legitimate decryption, the proposed architecture shifts the security paradigm from absolute prevention to accountable traceability. Watermarking, fingerprinting, access-event logging, and Merkle-proof validation collectively enable forensic attribution of unauthorized redistribution.

However, these benefits introduce trade-offs. First, blockchain transparency enhances auditability but may expose limited metadata signals. Even without storing medical content on-chain, access timestamps and ownership transfers could reveal contextual research patterns. High-sensitivity deployments may therefore require batching strategies, privacy-preserving logging, or zero-knowledge–based access proofs.

Second, the framework introduces governance and operational complexity. Smart contracts reduce discretionary ambiguity but increase reliance on correct contract design, secure key management, and reliable monitoring infrastructure. Formal verification and independent auditing are essential to mitigate smart contract vulnerabilities.

Third, the system assumes partial trust in auxiliary components such as key management services and custodial recovery mechanisms. While MPC-based designs reduce unilateral compromise risks, they do not eliminate operational dependencies.

Fourth, legal enforceability of NFT-based ownership remains jurisdiction-dependent. Cryptographic proof of control does not automatically equate to statutory recognition of ownership rights. Institutional integration with contractual agreements and regulatory frameworks is therefore required.

Fifth, long-term sustainability considerations, including decentralized storage persistence incentives and transaction costs, must be evaluated for large-scale production environments.

Importantly, this framework does not advocate the speculative tokenization of medical data. Instead, it positions NFTs as programmable control instruments for regulated data stewardship. Compared with conventional governance models, the architecture uniquely combines:

1. Cryptographically verifiable ownership and integrity binding.
2. Programmable, revocable, fine-grained access control.
3. Post-distribution forensic attribution mechanisms.

Together, these properties suggest a unified technical model in which ownership, access, and accountability are integrated rather than administratively fragmented.

Empirical validation through institutional pilot deployments remains necessary to assess usability, governance overhead, and cross-jurisdictional compliance dynamics. Nevertheless, the proposed framework establishes a technically coherent foundation for accountable medical data lifecycle management.

7. Conclusion and Future Work

This paper introduced a comprehensive NFT-based architecture for preserving ownership, integrity, and controlled access to sensitive medical datasets. By integrating blockchain infrastructure, decentralized encrypted storage, Merkle-root commitments, programmable smart contracts, and custodial key management, the framework addresses persistent challenges in medical data governance, including unauthorized redistribution, limited provenance traceability, and lack of post-access accountability.

Unlike conventional repositories that provide static ownership assertions, the proposed approach treats NFTs as active governance instruments. Each dataset NFT cryptographically binds encrypted medical data to its rightful owner, while smart contracts regulate programmable, revocable, and auditable access workflows. The incorporation of ephemeral decryption keys, watermark-based attribution, and forensic evidence generation mechanisms strengthens accountability beyond traditional data-sharing agreements.

Theoretical analysis demonstrates that the hybrid on-chain/off-chain architecture is computationally efficient and economically viable when deployed on scalable blockchain infrastructures. Strong integrity and ownership guarantees can therefore be achieved without prohibitive performance or financial overhead.

The principal contributions of this study are:

1. Novel NFT-based framework for cryptographic ownership preservation of medical datasets.
2. Merkle-root binding mechanism linking NFTs to encrypted off-chain medical data.
3. Programmable smart contract model enabling fine-grained, revocable access control.
4. Integrated leakage detection and forensic evidence generation mechanism.
5. Custodial key recovery architecture tailored for regulated medical environments.

Future research should focus on enhancing privacy, interoperability, and institutional readiness.

Privacy-preserving computation techniques, including secure multi-party computation, trusted execution environments, and zero-knowledge proofs could enable analytical workflows without direct plaintext exposure. Zero-knowledge access proofs may further reduce identity disclosure risks in transparent logging environments.

Interoperability can be strengthened by aligning NFT metadata schemas with established healthcare standards such as HL7 FHIR and DICOM profiles, enabling seamless integration between blockchain governance layers and clinical information systems.

Smart contract extensions may incorporate programmable licensing terms, usage-based royalties, and incentive mechanisms for data contributors. Cross-chain NFT architectures could enable dataset portability while preserving verifiable ownership histories.

Institutional trust would benefit from formal smart contract verification, systematic security audits, and regulatory compliance assessments aligned with GDPR and HIPAA. Advanced key management approaches, including threshold cryptography and dynamic key fragmentation, may further reduce cryptographic single points of failure.

Finally, pilot deployments within hospitals, research institutions, and ethics committees are essential to evaluate governance workflows, usability, interoperability, and real-world adoption constraints. Empirical evaluation will determine the operational viability of this architecture at scale.

References

[1] Conduah, A. K., Ofoe, S., & Siaw-Marfo, D. (2025). Data privacy in healthcare: Global challenges and solutions. *Digital Health, 11*, 20552076251343959. https://doi.org/10.1177/20552076251343959

[2] Andrew, J., Isravel, D. P., Sagayam, K. M., Bhushan, B., Sei, Y., & Eunice, J. (2023). Blockchain for healthcare systems: Architecture, security challenges, trends and future directions. *Journal of Network and Computer Applications, 215*, 103633. https://doi.org/10.1016/j.jnca.2023.103633

[3] Taloba, A. I., & Rayan, A. (2025). A privacy preserving medical data management framework using blockchain enabled encrypted role based access control. *Scientific Reports, 15*, 43864. https://doi.org/10.1038/s41598-025-30916-3

[4] Ullah, A., Ullah, Z., Rizvi, S. S., et al. (2025). Toward blockchain based electronic health record management with fine grained attribute based encryption and decentralized storage mechanisms. *Scientific Reports, 15*, 34542. https://doi.org/10.1038/s41598-025-17875-5

[5] Psarra, E., Apostolou, D., Verginadis, Y., Patiniotakis, I., & Mentzas, G. (2024). Permissioned blockchain network for proactive access control to electronic health records. *BMC Medical Informatics and Decision Making, 24*(1), 303. https://doi.org/10.1186/s12911-024-02708-8

[6] Li, K., Lohachab, A., Dumontier, M., & Urovi, V. (2025). Privacy preservation in blockchain-based healthcare data sharing: A systematic review. *Peer-to-Peer Networking and Applications, 18*(6), 302. https://doi.org/10.1007/s12083-025-02148-9

[7] Ni, J., Fang, G., Zhao, Y., Ren, J., Chen, L., & Ren, Y. (2024). Distributed group key management based on blockchain. *Electronics, 13*(11), 2216. https://doi.org/10.3390/electronics13112216

[8] Kaisto, J., Juutilainen, T., & Kauranen, J. (2024). Non-fungible tokens, tokenization, and ownership. *Computer Law & Security Review, 54*, 105996. https://doi.org/10.1016/j.clsr.2024.105996

[9] Ramirez Lopez, L. J., & Morillo Ledezma, G. G. (2025). Employing blockchain, NFTs, and digital certificates for unparalleled authenticity and data protection in source code: A systematic review. *Computers, 14*(4), 131. https://doi.org/10.3390/computers14040131

[10] Singh, P., Sagar, S., Singh, S., et al. (2024). Blockchain-enabled verification of medical records using soul-bound tokens and cloud computing. *Scientific Reports, 14*, 24830. https://doi.org/10.1038/s41598-024-75708-3

[11] Le, H. V. A., Nguyen, Q. D. N., Tadashi, N., & Tran, T. H. (2025). Blockchain-based decentralized identity management system with AI and Merkle trees. *Computers, 14*(7), 289. https://doi.org/10.3390/computers14070289

[12] Ezz, M., Alaerjan, A. S., & Mostafa, A. M. (2025). Ethical AI in healthcare: Integrating zero-knowledge proofs and smart contracts for transparent data governance. *Bioengineering (Basel), 12*(11), 1236. https://doi.org/10.3390/bioengineering12111236

[13] Petrovskiy, E. D. (2025). Dense longitudinal single-subject multimodal MRI dataset with self-administered acquisition. *OpenNeuro Dataset*. https://doi.org/10.18112/openneuro.ds006772.v1.0.5

[14] Liu, X. Y., Wang, W. L., Liu, M., Chen, M. Y., Pereira, T., Doda, D. Y., … Chen, X. (2025). Recent applications of EEG-based brain-computer-interface in the medical field. *Military Medical Research, 12*(1), 14. https://doi.org/10.1186/s40779-025-00598-z
[15] Xu, M., Ouyang, Y., & Yuan, Z. (2023). Deep learning aided neuroimaging and brain regulation. *Sensors, 23*(11), 4993. https://doi.org/10.3390/s23114993
[16] Klonoff, D. C., Scheideman, A. F., Shao, M. M., Ho, C. N., Ayers, A. T., Jendle, J., & Wykowski, H. G. (2026). The need for clear medical data ownership laws. *Journal of Translational Medicine, 24*(1), 5. https://doi.org/10.1186/s12967-025-07486-z
[17] Alvarez-Romero, C., Martínez-García, A., & Bernabeu-Wittel, M. (2023). Health data hubs: An analysis of existing data governance features for research. *Health Research Policy and Systems, 21*(1), 70. https://doi.org/10.1186/s12961-023-01026-1
[18] Hamdaoui, I., Makkaoui, K. E., El Allali, Z., & Maleh, Y. (2026). Securing trustworthy big data: Comprehensive overview of challenges and enhanced security with machine learning and blockchain. *Information Security Journal: A Global Perspective*, 1–35. https://doi.org/10.1080/19393555.2026.2616295
[19] Bonomi, L., Huang, Y., & Ohno-Machado, L. (2020). Privacy challenges and research opportunities for genomic data sharing. *Nature Genetics, 52*, 646–654. https://doi.org/10.1038/s41588-020-0651-0
[20] Beck, R., Mueller-Bloch, C., & King, J. L. (2018). Governance in the blockchain economy: A framework and research agenda. *Journal of the Association for Information Systems*, 1020–1034. https://doi.org/10.17705/1jais.00518
[21] Rokni, S. F., & Yavari, M. (2025). Exploring non-fungible tokens: A bibliometric analysis and future research opportunities. *Journal of Economic Surveys*. https://doi.org/10.1111/joes.70022
[22] Nunes, T., da Cunha, P. R., de Abreu, J. M., & Duarte, J., Corte-Real, A. (2024). Non-fungible tokens (NFTs) in healthcare: A systematic review. *International Journal of Environmental Research and Public Health, 21*(8), 965. https://doi.org/10.3390/ijerph21080965
[23] Ullah, A., Ullah, Z., Rizvi, S. S., & Gul, L., Kwon, S. J. (2025). Toward blockchain based electronic health record management with fine grained attribute based encryption and decentralized storage mechanisms. *Scientific Reports, 15*(1), 34542. https://doi.org/10.1038/s41598-025-17875-5
[24] Kumar, N., Kumar, K., Aeron, A., & Verre, F. (2025). Blockchain technology in supply chain management: Innovations, applications, and challenges. *Telematics and Informatics Reports, 18*, 100204. https://doi.org/10.1016/j.teler.2025.100204
[25] Alatawi, M. N. (2025). Blockchain-driven smart contracts for advanced authorization and authentication in cloud security. *Electronics, 14*(15), 3104. https://doi.org/10.3390/electronics14153104
[26] Razaque, A., Shaldanbayeva, N., Alotaibi, B., Alotaibi, M., Murat, A., & Alotaibi, A. (2022). Big data handling approach for unauthorized cloud computing access. *Electronics, 11*(1), 137. https://doi.org/10.3390/electronics11010137
[27] Hwang, Y. W., & Lee, I. Y. (2020). A study on CP-ABE-based medical data sharing system with key abuse prevention and verifiable outsourcing in the IoMT environment. *Sensors (Basel), 20*(17), 4934. https://doi.org/10.3390/s20174934
[28] Ahmed, A., Shahzad, A., Naseem, A., Ali, S., & Ahmad, I. (2025). Evaluating the effectiveness of data governance frameworks in ensuring security and privacy of healthcare data: A quantitative analysis of ISO standards, GDPR, and HIPAA in blockchain technology. *PLoS One, 20*(5), e0324285. https://doi.org/10.1371/journal.pone.0324285
[29] Golightly, L., Modesti, P., Garcia, R., & Chang, V. (2023). Securing distributed systems: A survey on access control techniques for cloud, blockchain, IoT and SDN. *Cyber Security and Applications, 1*, 100015. https://doi.org/10.1016/j.csa.2023.100015
[30] Bendayan, S., Cohen, Y., Bendayan, J., Windisch, S., & Afilalo, J. (2024). Nonfungible tokens in cardiovascular medicine. *Canadian Journal of Cardiology, 40*(10), 1959–1964. https://doi.org/10.1016/j.cjca.2024.07.010
[31] He, Y., Zhou, Z., Pan, Y., Chong, F., Wu, B., Xiao, K., & Li, H. (2024). Review of data security within energy blockchain: A comprehensive analysis of storage, management, and utilization. *High-Confidence Computing, 4*(3), 100233. https://doi.org/10.1016/j.hcc.2024.100233
[32] Li, X., & Cong, Y. (2024). Exploring barriers and ethical challenges to medical data sharing: Perspectives from Chinese researchers. *BMC Medical Ethics, 25*(1), 132. https://doi.org/10.1186/s12910-024-01135-8
[33] Esmaeilzadeh, P. (2023). Evolution of health information sharing between health care organizations: Potential of nonfungible tokens. *Interactive Journal of Medical Research, 12*, e42685. https://doi.org/10.2196/42685

[34] Liu, Z., Shi, Z., Wang, W., Kong, R., Fu, D., & Qiu, J. (2025). Research on data ownership and controllable sharing schemes in the process of logistics data flow. *Electronics, 14*(9), 1714. https://doi.org/10.3390/electronics14091714
[35] Hanif, M., Munir, E. U., Rehan, M. M., Ahmad, S. G., Khan, I., & Setchi, R. (2025). Tiered blockchain framework: A secure, trustworthy, and cost-efficient solution for the digital rights protection. *Blockchain: Research and Applications, 6*(4), 100308. https://doi.org/10.1016/j.bcra.2025.100308
[36] Habib, G., Sharma, S., Ibrahim, S., Ahmad, I., Qureshi, S., & Ishfaq, M. (2022). Blockchain technology: Benefits, challenges, applications, and integration of blockchain technology with cloud computing. *Future Internet, 14*(11), 341. https://doi.org/10.3390/fi14110341
[37] Singh, J., Bharany, S., Rani, S., et al. (2025). A systematic review of blockchain, AI, and cloud integration for secure digital ecosystems. *International Journal of Networked and Distributed Computing, 13*, 28. https://doi.org/10.1007/s44227-025-00072-1
[38] Torab-Miandoab, A., Samad-Soltani, T., Jodati, A., Akbarzadeh, F., & Rezaei-Hachesu, P. (2024). A unified component-based data-driven framework to support interoperability in healthcare systems. *Heliyon, 10*(15), e35036. https://doi.org/10.1016/j.heliyon.2024.e35036
[39] Khalil, U., Malik, O. A., Uddin, M., & Chen, C.-L. (2022). A comparative analysis on blockchain versus centralized authentication architectures for IoT-enabled smart devices in smart cities: A comprehensive review, recent advances, and future research directions. *Sensors, 22*(14), 5168. https://doi.org/10.3390/s22145168
[40] Leite, E., & Ingstrup, M. B. (2022). Individual strategies as interaction modes for handling institutional logic diversity over time: A case study on a public-private collaboration project. *Industrial Marketing Management, 107*, 266–275. https://doi.org/10.1016/j.indmarman.2022.09.019
[41] Chao, K., Sarker, M. N. I., Ali, I., Firdaus, R. B. R., Azman, A., & Shaed, M. M. (2023). Big data-driven public health policy making: Potential for the healthcare industry. *Heliyon, 9*(9), e19681. https://doi.org/10.1016/j.heliyon.2023.e19681
[42] van Daalen, O. L. (2023). The right to encryption: Privacy as preventing unlawful access. *Computer Law & Security Review, 49*, 105804. https://doi.org/10.1016/j.clsr.2023.105804
[43] Martínez-Ballesté, A., Batista, E., Figueroa, E., et al. (2026). Security and privacy in complex, context-aware systems: A comprehensive framework. *International Journal of Information Security, 25*, 57. https://doi.org/10.1007/s10207-026-01228-y
[44] Razi, Q., Devrani, A., Abhyankar, H., Chalapathi, G. S. S., Hassija, V., & Guizani, M. (2024). Non-fungible tokens (NFTs)—Survey of current applications, evolution, and future directions. *IEEE Open Journal of the Communications Society, 5*, 2765–2791. https://doi.org/10.1109/OJCOMS.2023.3343926
[45] Kuznetsov, O., Frontoni, E., Kuznetsova, K., Shevchuk, R., & Karpinski, M. (2024). NFT technology for enhanced global digital registers: A novel approach to tokenization. *Future Internet, 16*(7), 252. https://doi.org/10.3390/fi16070252
[46] Davies, J. (2024). Enhanced scalability and privacy for blockchain data using Merklized transactions. *Frontiers in Blockchain, 6*, 1222614. https://doi.org/10.3389/fbloc.2023.1222614
[47] Kuznetsov, O., Rusnak, A., Yezhov, A., Kuznetsova, K., Kanonik, D., & Domin, O. (2024). Evaluating the security of Merkle trees: An analysis of data falsification probabilities. *Cryptography, 8*(3), 33. https://doi.org/10.3390/cryptography8030033

Statements

Data Availability

The dataset used in this study is publicly available through the OpenNeuro repository under a CC0 license (accession number DS007328). No new medical data were generated for this work.

Code availability

The prototype implementation of the proposed framework is publicly available at: https://github.com/hodanaseri/NFT-based-medical-dataset-ownership

Ethics approval and consent to participate

This study did not involve direct interaction with human participants. All data used were obtained from an existing publicly available repository (OpenNeuro DS007328), which has been fully anonymized by its or iginal curators. No new medical data were collected or generated for this work.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable

Funding

Not applicable

Author information

Hoda Naseri was responsible for the conceptualization and core methodology of this research. Seyed Mohammad Mirhosseini collaborated in investigation. Ali A. Safaei supervised the research process and provided critical revisions.

Acknowledgements

Not applicable