International Journal of Computer Applications Technology and Research
Volume 11–Issue 12, 711 - 723, 2022, ISSN:-2319–8656
DOI:10.7753/IJCATR1112.1031

Developing Standardized Metadata Protocols Enabling Transparent Provenance Tracking for AI-Created Media Within Federal Intellectual Property Regulatory Systems Nationwide

Precious Mathias Omogiate
Legal Associate,
Chief A.B Thomas Legal
Practitioners and Arbitrators,
Lagos, Nigeria

Abstract: The rapid proliferation of artificial intelligence (AI) systems capable of generating images, audio, video, and written content has challenged the foundations of traditional intellectual property (IP) governance and evidence frameworks across federal regulatory systems. As AI-generated media becomes increasingly indistinguishable from human-authored works, questions of authorship, ownership, authenticity, and liability have become significantly more complex. Current federal IP regulations, originally built around clear human creative intent and manual documentation, lack standardized mechanisms for verifying the provenance and creative lineage of AI-derived outputs. This gap not only complicates copyright claims and patent defenses but also opens pathways for misinformation campaigns, fraudulent media manipulation, and digital marketplace disputes. To address these concerns, the development of standardized metadata protocols is emerging as a critical priority. Such protocols would embed persistent, machine-readable provenance markers capturing model origin, training data sources, transformation history, and authorship contributions into AI-created media at the point of generation. When implemented across federal IP workflows, these metadata structures would enable transparent chain-of-custody tracking and facilitate reliable audit, authentication, and enforcement procedures. The proposed framework requires interoperability across commercial AI tools, regulatory databases, and digital asset registries, supported by secure hashing, digital signatures, and tamper-evident storage. By institutionalizing standardized metadata protocols nationwide, federal systems can better safeguard creative rights, uphold marketplace integrity, and promote responsible innovation, ensuring that AI-mediated creativity remains verifiable, accountable, and legally protected.

Keywords: AI provenance, metadata standardization, intellectual property regulation, digital authenticity, federal compliance systems, generative media governance.

1. INTRODUCTION

1.1 Context: Rise of AI-Generated Media

AI-driven content generation expanded rapidly with the advancement of neural networks capable of producing text, images, video, and audio in ways that closely mimic human creative patterns [1]. These systems draw from large-scale datasets and computational learning loops to construct outputs that appear intentional, expressive, and stylistically coherent, even without direct human authorship [2]. As media workflows became increasingly automated, creative tasks that once required specialized human labor illustration, compositing, voice synthesis, and narrative drafting became partially or fully automated across journalism, entertainment, marketing, and software documentation [3]. The resulting content ecosystems blurred distinctions between original creation, algorithmic transformation, and statistical recomposition, shifting value from human technique to system configuration and dataset curation [4]. Organizations integrated these generative capabilities into existing production pipelines, accelerating turnaround times and reducing costs while expanding volume and variation of outputs [5]. However, this diffusion occurred faster than the establishment of ethical, legal, or professional guardrails. The challenges that emerged were not limited to authenticity or verification; they included uncertainty around creative ownership, authorship recognition, and economic reward structures for both developers and creative workers [6]. These dynamics formed the foundational environment in which questions of provenance and attribution gained heightened significance.

1.2 Breakdown in Traditional Authorship Models

Historically, authorship has been associated with identifiable human intention, expressive agency, and traceable creative lineage [7]. Legal and cultural frameworks linked originality to the skilled application of personal knowledge and creative judgment, forming the basis for copyright and moral rights protections [4]. Yet generative AI systems complicate this foundation. These models draw from vast and often opaque training datasets, containing fragments of stylistic patterns, compositional frameworks, and representational cues originating from many unknown contributors [8]. The output produced may not be directly traceable to any singular creative source, nor can it be accurately described as purely authored by a human operator who merely supplies prompts or high-level input [2]. This dissolves the conventional link between creator identity and creative product, weakening the reliability of authorship as a legal and economic category [9]. As attribution becomes ambiguous, compensation and recognition structures for creative labor become unstable,

International Journal of Computer Applications Technology and Research
Volume 11–Issue 12, 711 - 723, 2022, ISSN:-2319–8656
DOI:10.7753/IJCATR1112.1031

driving disputes not only among creators, but among platform developers, rights holders, and regulatory institutions [7]. Without mechanisms to record how AI systems derive, transform, and assemble media, authorship definitions risk becoming functionally hollow, threatening the economic incentives that support original creative practice.

1.3 Federal Regulatory Urgency for Provenance Frameworks

In response to the erosion of clear authorship boundaries, federal institutions increasingly require mechanisms that preserve the traceability of creative processes and ownership claims [5]. Provenance frameworks provide a structured method for documenting origin, transformation steps, model influence, and human versus machine contribution within media artifacts [3]. Such metadata must be persistent, machine-readable, and secure against alteration to serve as credible evidence in intellectual property administration, legal arbitration, and commercial licensing environments [9]. The absence of standardized provenance protocols risks allowing unverifiable authorship assertions, fraudulent claims over AI-generated works, and disputes that undermine regulatory consistency and judicial fairness [6]. Moreover, industries relying on trusted communication, such as education, journalism, policy, and healthcare, require stable authentication signals to prevent misinformation, manipulation, and misattributed expertise [8]. Federal regulatory urgency therefore stems from the need to safeguard both market integrity and the conditions of cultural production [4]. Without a coherent nationwide approach, individual organizations may implement incompatible or proprietary attribution mechanisms, fragmenting compliance landscapes and obstructing interoperability across agencies and sectors [7]. Establishing shared metadata standards is essential not only to clarify creative ownership but to maintain transparency, accountability, and public trust in AI-mediated communication ecosystems.

2. CONCEPTUAL AND TECHNICAL BACKGROUND

2.1 Defining Digital Provenance and Metadata Standards

Digital provenance refers to the structured record of origin, authorship, transformation, and ownership lineage associated with a media artifact across its lifecycle [12]. It represents not only where a piece of content came from, but also how it has been altered, by whom, and under what conditions. Within digital creative ecosystems, provenance relies on metadata, which functions as descriptive and evidentiary information attached to or embedded within media files [9]. Metadata standards establish the schemas, formats, and protocols through which this information is stored, interpreted, and exchanged across platforms, applications, and regulatory systems [14]. Traditional metadata practices were developed for archival governance, broadcast asset management, and copyright cataloging, where human authorship and controlled production pipelines could be assumed [10]. However, when applied to generative AI media, provenance must account for model versioning, dataset sources, prompt structures, algorithmic decision layers, and human contribution differentials [16]. Without standardized metadata fields that capture these dimensions, provenance becomes ambiguous or incomplete. Further, provenance integrity requires resilience against tampering, loss, or dissociation from the underlying artifact over time [8]. Accordingly, the emerging metadata challenge is not merely descriptive but infrastructural: systems must support persistent, interoperable, verifiable provenance that remains intact across production, storage, publication, and redistribution environments [17]. This establishes the conceptual foundation for examining how AI model outputs complicate provenance tracking at scale.

2.2 Overview of Generative AI Model Architectures and Output Pipelines

Generative AI models operate by learning statistical relationships across large datasets and using these learned patterns to produce new content that resembles the data on which they were trained [11]. Architectures such as generative adversarial networks (GANs), transformer-based language models, diffusion models, and variational autoencoders each employ different computational strategies to synthesize outputs across modalities including text, image, audio, and video [13]. In most production workflows, generative media passes through multiple stages: model selection and configuration, prompt or input specification, inference generation, output refinement, and export to downstream editing or distribution environments [15]. Each stage introduces potential human or algorithmic influence, contributing to the complexity of attributing authorship and tracking decision lineage. For example, prompt engineering may guide thematic or stylistic direction, while post-generation editing may mask or obscure signals of synthetic origin [9]. Additionally, outputs may be automatically transcoded, compressed, or merged with other assets, each process potentially stripping, modifying, or overwriting provenance metadata [14]. These activities occur within a distributed pipeline involving users, platform interfaces, content delivery networks, and third-party applications.

Figure 1. High-Level Process Flow of Generative Media Creation and Distribution

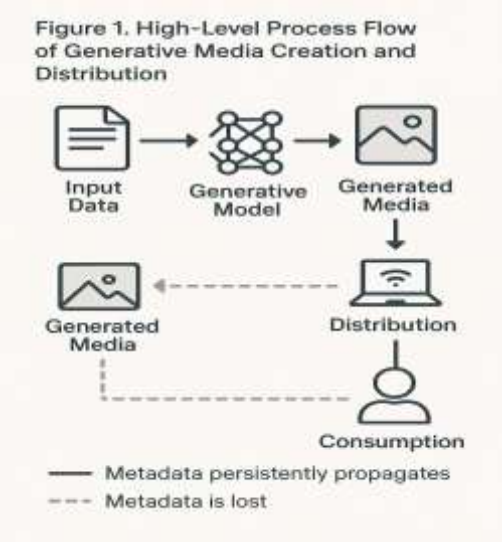


Figure 1 illustrates a high-level process flow of generative media creation and distribution pipelines, highlighting nodes where metadata may either propagate persistently or be lost.

Given the complexity and heterogeneity of pipelines, provenance cannot rely solely on user reporting or manual ledgering [12]. Instead, metadata capture must be embedded at generation, enforced throughout the pipeline, and verifiable independently of platform claims [16]. Understanding this pipeline architecture is essential for recognizing where provenance loss can occur and where automated audit mechanisms must intervene.

2.3 Existing Metadata Approaches: Strengths and Gaps

Current metadata systems were not designed to accommodate the opacity and variability of algorithmic creativity. Standard formats such as EXIF for images, ID3 tags for audio, and XMP sidecar schemas for multimedia allow descriptive metadata to accompany files, but these formats rely on voluntary attribution and can be easily removed, altered, or overwritten during file editing and transfer [8]. Some emerging platforms implement proprietary watermarking or invisible tagging techniques to identify AI-generated assets; however, these approaches lack interoperability and are not universally verifiable across heterogeneous toolchains [13]. Cryptographic hashing and blockchain-based asset registration offer more robust immutability guarantees but require consistent adoption across decentralized producers, which remains unlikely without regulatory or industry standardization incentives [17].

One of the most significant gaps lies in the failure to record the computational and dataset lineage of outputs. While some enterprise-oriented AI development platforms support model version tracking, dataset documentation, and training pipeline registries, this provenance information often remains internal and inaccessible to external auditors, users, or legal bodies [15]. Furthermore, many generative systems rely on training datasets derived from public web corpora or mixed-license archives, where the original creators are not identified or consent is uncertain [11]. This creates layered ambiguity: neither the provenance of input materials nor the contribution ratio of human vs. algorithmic labor is reliably traceable.

Attempts to supplement metadata with forensic analysis such as statistical artifact detection or stylistic pattern-matching are limited in reliability and vulnerable to adversarial manipulation [9]. Without standards enforcing persistent, tamper-resistant provenance identifiers at generation, supplemented by pipeline-wide metadata preservation rules, existing approaches cannot meet evidentiary thresholds required for legal adjudication, archival authentication, or professional attribution systems [14]. Thus, the need for a unified, enforceable, and interoperable provenance standard arises not from inadequacy of metadata formats themselves, but from their inability to maintain integrity, completeness, and authority under conditions of algorithmically mediated media production [12].

3. REGULATORY LANDSCAPE AND POLICY GAPS

3.1 Current Federal IP Protections Related to Digital Media

Federal intellectual property protections historically rely on the assumption that creative works originate from identifiable human authors exercising intentional and original expression [18]. Copyright law provides exclusive rights to reproduce, distribute, and derive value from such works, contingent upon a demonstrable connection between the creator and the creative product. In digital contexts, these protections extend to software, audiovisual content, textual compositions, and visual artworks, provided they meet requirements for minimal originality and fixation in a tangible medium [21]. Enforcement mechanisms rely on documented authorship claims, licensing records, and chain-of-title evidence maintained through contracts, registries, and metadata systems [16]. However, digital media workflows increasingly involve automated editing tools, algorithmic enhancement, and collaborative production environments where human creative control is partial or distributed [19]. The law continues to treat digital outputs under the same authorship principles applied to traditional media, presuming that human creators retain meaningful creative influence. Yet generative AI introduces content that may not originate from direct human intention in any conventional sense [15]. Federal agencies acknowledge the existence of machine-generated media but lack a unified framework specifying how authorship, ownership, and derivative rights apply when algorithms contribute substantial creative content [24]. As a result, existing IP protections remain structurally grounded in human-centered models that do not map cleanly onto AI-mediated production ecosystems [22].

3.2 Limitations in Statutory Definitions of Authorship for Algorithmic Creation

The statutory definition of authorship presumes that creative works are the product of human conceptual guidance, judgment, and expressive decision-making [20]. This requirement becomes problematic when generative AI models produce media outputs autonomously or with highly abstract forms of human prompting. Traditional authorship doctrine does not clearly distinguish between human-originated creative labor and machine-mediated synthesis, nor does it specify thresholds of human involvement necessary to establish ownership [17]. For example, a user who inputs a brief textual prompt may not exert creative control significant enough to satisfy originality requirements, yet the model itself cannot legally hold authorship because it lacks legal personhood and intent [23]. This creates a definitional void. In some contexts, authorship has been attributed to the human operator on the basis of initiating the generative process, while in others, courts and administrative bodies have ruled that works lacking identifiable human expression cannot be copyrighted at all [15]. The result is inconsistent and unpredictable classification of AI-generated works within

existing legal frameworks [19]. Furthermore, because AI models learn from datasets containing pre-existing creative works, questions arise regarding derivative use, contributory authorship, and whether model outputs indirectly embed or recombine protected expression from unknown contributors [24]. Without statutory clarification, disputes surrounding AI-mediated authorship risk escalating into protracted litigation and contradictory administrative rulings [22].

3.3 Risks of Non-Standardized Provenance in Federal Legal Proceedings

The absence of standardized provenance mechanisms significantly undermines the evidentiary stability of authorship and ownership claims in federal legal contexts [16]. When media artifacts lack verifiable metadata recording model identity, training lineage, prompts, editing history, and human contribution points, claims of authorship rely primarily on assertions rather than demonstrable proof [21]. This erodes the reliability of copyright registrations, licensing contracts, and infringement claims, particularly where multiple parties may assert rights based on indirect contribution or platform-level authorship claims [18]. Additionally, without persistent provenance, courts face difficulty determining whether a disputed work is original, derived, or synthetically recomposed from copyrighted material included in training datasets [23]. The potential for falsified authorship claims, concealed AI involvement, or untraceable creative lineage increases the likelihood of legal misattribution and inconsistent judicial outcomes [20]. Moreover, the lack of interoperable provenance standards forces agencies, litigants, and auditors to rely on platform-provided disclosures, which may be incomplete or strategically curated [24]. Accordingly, non-standardized provenance not only complicates enforcement but threatens the credibility and functionality of federal IP adjudication systems [17].

4. PROBLEM STATEMENT: CHALLENGES IN TRANSPARENT PROVENANCE TRACKING

4.1 Difficulty Tracing Training Data and Model Influence

Tracing the origins of content generated by AI models is challenging because these systems are trained on vast corpora composed of heterogeneous sources, including public repositories, licensed datasets, scraped web archives, and user-generated materials [24]. The training process abstracts patterns from these inputs rather than storing explicit copies, making it difficult to determine whether a specific output reflects transformative synthesis or recognizable derivation from a prior work [27]. In many cases, the provenance of the underlying training data is undocumented or partially documented due to historical collection practices prioritizing scale over attribution integrity [22]. Additionally, model updates, fine-tuning, and transfer learning further obscure lineage by blending knowledge across multiple training stages and contributing entities [28]. The absence of systematic tracking mechanisms prevents auditors from establishing how particular creative influences contribute to a final output, complicating claims regarding originality, authorship, and rightful ownership [25]. Without transparent tracking of dataset composition and model decision pathways, attribution remains ambiguous not only for legal purposes, but also for ethical and commercial contexts where recognition and compensation are at stake [29]. As a result, any framework for provenance must confront not only output verification, but the traceability of learning histories embedded within model architectures [30].

4.2 Lack of Persistent, Interoperable Metadata Across Platforms

Even when metadata describing authorship or production processes is generated, the lack of interoperability among platforms leads to frequent loss, alteration, or fragmentation of provenance information [23]. Different software environments, cloud infrastructures, creative tools, and publishing systems follow distinct metadata schemas and serialization formats, meaning provenance fields may not survive conversion, export, or editing operations [26]. In many cases, metadata is stored in optional or fragile layers such as sidecar files or application-specific attribute fields that can be easily stripped during compression, resizing, transcoding, or social media distribution [22]. Proprietary generative content platforms may embed identifying information, but these solutions rarely align with open standards and cannot consistently interoperate across decentralized production ecosystems [28]. Without persistent linkage between the digital artifact and its provenance record, authorship claims cannot be reliably verified when works are shared, modified, or redistributed [30]. Moreover, metadata that lacks cryptographic integrity or tamper resistance cannot function as trustworthy evidence in regulatory or judicial settings [24]. The absence of universal, authoritative provenance formats undermines cross-platform enforcement and prevents regulators from establishing consistent expectations for disclosure compliance [29]. The result is a fragmented provenance environment in which authenticity depends on platform-level policy rather than standardized infrastructure that ensures continuity and verification across the media lifecycle [27]. Addressing this lack of interoperability is therefore a fundamental requirement for national provenance frameworks capable of supporting legal, archival, and creative accountability.

4.3 Vulnerabilities to Manipulation, Forgery, and Omission

Provenance metadata can be intentionally removed, falsified, or replaced, creating opportunities for fraud, misrepresentation, and misattribution in creative markets [25]. Unscrupulous actors may erase or modify metadata to obscure the involvement of generative models, allowing AI outputs to be passed off as solely human-authored or as belonging to different rights holders [22]. This vulnerability is particularly concerning for industries that rely on verified originality or professional credentialing, where the misrepresentation of

International Journal of Computer Applications Technology and Research
Volume 11–Issue 12, 711 - 723, 2022, ISSN:-2319–8656
DOI:10.7753/IJCATR1112.1031

creative labor can distort compensation systems, undermine trust, and facilitate plagiarism or market manipulation [30]. Additionally, metadata may be omitted not through malicious intent but through ordinary technical processes, such as when editing tools default to saving files without embedded attribution fields [23]. The absence of tamper-evident controls leaves auditors and legal authorities dependent on platform disclosures, which may be incomplete or unverifiable [28].

Table 1 provides a comparative matrix illustrating the limitations of current provenance-tracking tools, highlighting weaknesses in persistence, verifiability, cross-platform compatibility, and adversarial resistance.

Without enforced metadata retention and validation mechanisms, provenance cannot guarantee authenticity or authorship integrity [26]. These vulnerabilities reveal that provenance must be safeguarded not only through descriptive metadata fields but through embedded integrity protections, cryptographic binding, and infrastructure-level enforcement that travels with the digital artifact and remains intact across iterative transformation and distribution environments [29].

Table 1. Comparative Matrix of Current Provenance-Tracking Tools and Their Deficiencies

Tool / Approach	Persistence of Metadata Across Editing & Distribution	Verifiability of Authorship and Model Influence	Cross-Platform Interoperability	Resistance to Manipulation / Forgery
EXIF Metadata (Images / Media Files)	Weak – removed easily when compressed, edited, or shared on social platforms	Very Low – provides no evidence of algorithmic or model-based creation	Low – inconsistent support across software and cloud platforms	Very Low – easily editable or erasable with basic tools
XMP Sidecar Files	Moderate – can persist if workflows preserve sidecar; lost frequently in distribution	Low – does not document training data, prompts, or model lineage	Low-Moderate – dependent on compliant applications; breaks when files are detached	Low – metadata lives outside the file and is easily replaced
Proprietary Platform Watermarking (e.g., internal AI service tagging)	Moderate – persists only within the same platform ecosystem	Moderate – verifies platform of origin, not training data influences or human contribution	Very Low – usually non-portable and unreadable outside vendor environment	Low-Moderate – can be spoofed, removed, or bypassed when content is exported or screen-captured
Invisible Signal or Fingerprint Embedding (e.g., pixel-level perturbations)	Variable – may survive resizing or format change, but fragile under recompression or editing	Low – indicates synthetic origin but does not explain authorship or dataset attribution	Very Low – no standardized detection infrastructure across tools	Low – adversarial attacks can mask or erase fingerprints
Cryptographic Hashing (File-Integrity Checks)	High – detects post-creation modification reliably	Moderate – confirms unchanged origin, but does not describe how the content was generated	High – platform-independent and algorithmically consistent	High for integrity; Low for attribution – cannot indicate authorship without additional metadata linkage
Blockchain / Distributed Ledger Provenance Logs	Very High – append-only records resistant to later alteration	High – can document model versioning, contributors, and timestamps when properly implemented	Moderate – requires compatible registry interfaces for cross-system integration	High – tamper-evident, but dependent on initial correctness of submitted metadata
Enterprise	High	High for	Very Low –	High

Tool / Approach	Persistence of Metadata Across Editing & Distribution	Verifiability of Authorship and Model Influence	Cross-Platform Interoperability	Resistance to Manipulation / Forgery
MLOps Model Cards / Dataset Documentation Systems	within organization; Low when assets leave internal environment	model lineage; Low for final media outputs unless paired with embedded metadata	typically proprietary and not shared across organizational boundaries	internally; Low externally due to lack of standardized validation layers

5. PROPOSED STANDARDIZED METADATA PROTOCOL FRAMEWORK

5.1 Architectural Principles for Metadata Standardization

A national metadata standard for AI-generated media must adopt architectural principles that promote durability, transparency, and verification across diverse technological and institutional environments [32]. The first principle is persistent binding between a media artifact and its provenance record: metadata must remain inseparable from the asset across editing, storage, and redistribution workflows [29]. This requires embedding metadata at the asset level rather than relying on external documentation files that can be easily lost or altered. The second principle is granularity. Provenance records must capture not only static authorship claims, but also detailed information about model configuration, training influences, prompt origin, and human involvement at each stage of content generation [35]. Without granular records, provenance becomes too generalized to support legal attribution or audit-quality traceability. Third, the system must adhere to machine-readability and automation-readiness, enabling compliance checks and verification to run at scale rather than relying on manual inspection [31]. Fourth, metadata standards must be tamper-evident, incorporating cryptographic markers so that unauthorized modification or removal can be detected, even when assets are repeatedly transcoded or transferred [34]. Finally, the architecture must support interoperability, ensuring consistent interpretation across platforms, proprietary systems, and jurisdictional boundaries [36]. Taken together, these principles establish a foundation for provenance infrastructures capable of sustaining authorship integrity and ownership transparency in generative media ecosystems, while also supporting institutional enforcement mechanisms and administrative review processes [30].

5.2 Core Metadata Field Specifications

A standardized metadata schema for AI-generated media must define explicit fields that encode key aspects of model behavior, user contribution, and content lineage [33]. The first essential field is Model Identifier, including model name, version number, and any fine-tuning fingerprints, enabling auditors to determine which computational architecture influenced the output [29]. The Training Dataset Identifier is equally critical, documenting dataset category, licensing status, and provenance of training materials to clarify whether derivative or transformative properties may apply in legal contexts [37]. The schema should also include Generation Timestamp, recorded in secure, standardized format to support chronological sequencing in chain-of-custody verification [32].

Next, the Contributor Identity Record must distinguish between human and algorithmic contributions. For human contributors, identifiers should reflect authorship roles (e.g., prompt author, post-processing editor). For machine contributions, identifiers should reference model execution modules or transformation layers [30]. Additional fields include Prompt or Input Reference, encoded in hashed form to balance transparency and privacy, and Transformation Log, documenting post-generation edits, filters, or recomposition steps [35]. The metadata must also support Usage and Licensing Declarations, providing clarity on allowed distribution, adaptation, and resale conditions [31].

To ensure reliability, each field requires standardized formats and controlled vocabularies so values are comparable across platforms and can be ingested into legal, archival, or audit systems [34]. Without harmonized field specifications, provenance tracking becomes inconsistent, reducing its evidentiary power and interoperability across systems that rely on lineage authentication [36].

5.3 Persistent and Tamper-Evident Encoding Techniques

Ensuring that metadata remains durable and verifiable requires tamper-evident encoding methods that secure provenance against alteration, erasure, or substitution [29]. Cryptographic hashing is a foundational technique, generating fixed-length signatures representing both the media artifact and its metadata record [32]. Any change to the content or metadata alters the hash, enabling automated detection of manipulation. Digital signatures expand upon this by binding hashes to authenticated entities, confirming the identity of contributors or registrars involved in generation, editing, or approval workflows [34].

In distributed environments, blockchain or distributed ledger systems can record signatures, timestamps, and identity attestations in an append-only format, preventing retroactive editing or deletion of provenance history [36]. These ledgers support decentralized verification, allowing regulators, publishers, and auditors to validate authenticity without relying on proprietary vendor records [30].

However, tamper-evident mechanisms must be embedded not only in storage layers but at generation time. Metadata should be captured and encrypted at the point of model inference, not appended later, to prevent post-hoc reconstruction that could obscure authorship pathways [35].

Figure 2 presents a layered architecture model illustrating how cryptographic binding integrates with model execution pipelines and registry systems to support persistent provenance across content lifecycles.

Without these encoding structures, provenance remains vulnerable to manipulation, weakening trust in authorship attribution claims and undermining legal and administrative evaluation processes [33].

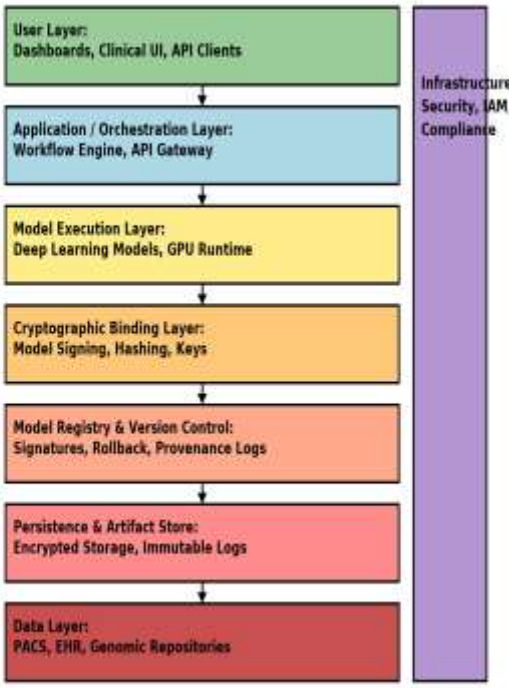


Figure 2. Layered Architecture Model for End-to-End Cryptographic Provenance in AI Model Pipelines

5.4 Interoperability Across Platforms, Agencies, and Media Formats

For provenance metadata to serve national regulatory functions, it must operate consistently across heterogeneous technological environments, including commercial platforms, public communication systems, archival repositories, and legal review infrastructures [31]. The foundation of interoperability lies in shared schema standards, ensuring that metadata fields retain meaning regardless of the application or storage environment in which they are encountered [29]. To support cross-agency collaboration, metadata must also align with administrative workflows used by copyright offices, evidentiary registries, and digital identity infrastructure frameworks [37].

Technical interoperability requires transport-independent embedding, meaning metadata persists whether a file is compressed, transcoded, or imported into a new editing system [36]. This can be achieved through container-level embedding with fallback sidecar synchronization to support legacy systems [30]. Application-level interoperability relies on API-based exchange protocols that allow provenance systems to verify metadata integrity without requiring full access to the media artifact itself [35].

Institutional interoperability demands procedural coordination among agencies that regulate media, authenticate authorship, or adjudicate disputes [33]. Without coordinated standards, agencies risk issuing contradictory rulings or evidentiary assessments, undermining trust in adjudication outcomes [32]. Interoperability ensures that provenance is not merely technically consistent, but also operationally authoritative, fulfilling its role in legal, archival, commercial, and cultural governance systems [34].

5.5 Governance Structure for National Implementation

A national provenance framework requires structured governance to ensure consistency, accountability, and adaptability over time [29]. Governance should be coordinated by a multi-stakeholder standards body composed of federal agencies, research institutions, creative labor organizations, and platform developers [31]. This body would define baseline metadata specifications, certify compliant toolchains, and maintain compatibility guidelines across software ecosystems [36].

Compliance enforcement must rely on mandatory adoption for systems participating in regulated distribution channels, supported by audit and certification mechanisms that verify metadata persistence and authenticity integrity [35]. The governance structure should also include dispute resolution procedures, enabling authors, platforms, and regulators to challenge or validate provenance assertions without requiring full judicial proceedings [33].

Finally, governance must support iterative revision to accommodate evolving model architectures, ethical norms, and media practices [37]. Sustained oversight ensures that provenance standards remain aligned with creative innovation while safeguarding authorship accountability and public trust [32].

6. IMPLEMENTATION AND INTEGRATION IN FEDERAL SYSTEMS

6.1 Integration with Copyright Offices, USPTO, Federal Archives

Integrating provenance metadata into federal intellectual property and archival infrastructures requires systematic alignment of technical workflows, evidentiary standards, and administrative review procedures [35]. Copyright offices have

long relied on self-declared authorship documentation supported by accompanying deposit materials; however, this model assumes that claimants can reliably assert creative origin and ownership. AI-generated media challenges this assumption by introducing layered contributions that may not be visible or traceable without embedded metadata linking outputs to models, prompts, and datasets [33]. To support reliable adjudication, registration systems must incorporate automated ingestion of provenance metadata at the point of submission, enabling examiners to verify declared authorship against algorithmic contribution records stored in authenticated provenance fields [38].

Similarly, integration with the U.S. Patent and Trademark Office (USPTO) requires explicit documentation of the extent to which algorithmic processes inform inventive steps or creative transformations [37]. Patent examination workflows could reference standardized provenance logs to differentiate between human-derived conceptual advances and machine-generated variations. Federal archives, which prioritize preservation and historical authenticity, require provenance to remain durable across format migrations, institutional transfers, and long-term storage cycles [39].

Achieving this integration necessitates shared schema definitions, secure metadata transport protocols, and validation services accessible across agencies [34]. Moreover, institutional adoption must include training for examiners and archivists to interpret metadata fields in the context of authorship and originality evaluation, ensuring that provenance does not merely exist as technical documentation, but functions as authoritative evidence supporting regulatory decisions [40].

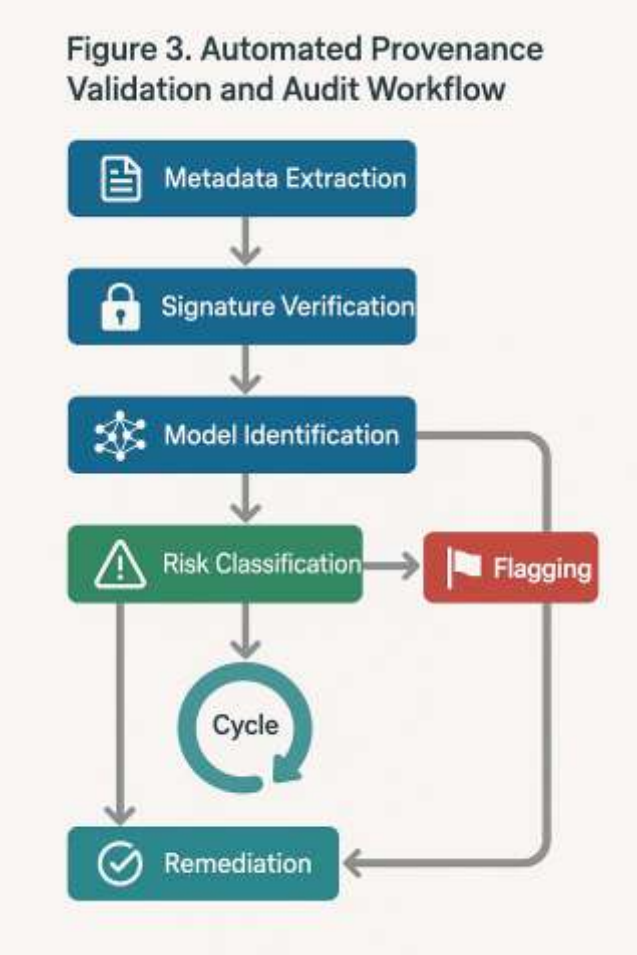
6.2 Automated Compliance and Authentication Workflows

Automated compliance workflows allow provenance validation to occur consistently, efficiently, and at scale across large volumes of AI-generated media [36]. Instead of relying on manual review or voluntary disclosures, automated systems can parse embedded metadata, verify cryptographic signatures, cross-check model identifiers, and assess contributor records through standardized audit rules [33]. These workflows begin at the point of content submission to a regulated distribution channel, such as a publishing platform, licensing marketplace, or rights registry. Metadata is extracted and checked against schema requirements and known registries of validated model versions and dataset identifiers [39]. If metadata fields are incomplete or fail verification, the system can automatically flag the artifact for secondary review or request correction before the content proceeds.

Authentication workflows incorporate forensic validation to detect undisclosed AI involvement. Statistical analysis, output fingerprint recognition, and probabilistic content classifiers can identify generation patterns indicative of specific model families, providing an additional verification layer when metadata is missing or compromised [35]. To prevent circumvention, automated workflows must be integrated not only in endpoint systems, but across editing suites, transcoding pipelines, and content management systems where metadata can be lost or overwritten [38].

Figure 3 depicts the automated provenance validation and audit workflow, showing metadata extraction, signature verification, model identification, risk classification, flagging, and remediation cycles.

Audit results must be logged in immutable records accessible to regulatory bodies and authorized stakeholders to support dispute resolution and legal evidence requirements [34]. For scalability, workflows must align with cloud-native orchestration environments, enabling continuous monitoring across high-volume and distributed production networks. Automated compliance ensures that provenance functions not only as a record of authorship but as an enforceable standard embedded into the infrastructure of media circulation [40].



6.3 Compatibility with Commercial AI Platforms and Media Production Software

Effective provenance tracking requires compatibility with widely used generative AI platforms, editing applications, and distribution environments [37]. Commercial AI providers vary in how they expose model identifiers, training data documentation, and usage logs. Some platforms include

International Journal of Computer Applications Technology and Research
Volume 11–Issue 12, 711 - 723, 2022, ISSN:-2319–8656
DOI:10.7753/IJCATR1112.1031

proprietary watermarking or attribution metadata, while others provide minimal or no disclosure [35]. This variability complicates integration, as provenance mechanisms must function regardless of vendor-specific implementation. Interoperability demands that metadata generation occur at the model inference layer and follow open schemas that remain interpretable when media is exported, recompressed, or modified in third-party software [33].

Media production environments introduce additional challenges because editing workflows often involve multiple applications that may not preserve embedded metadata. Production software, including image editors, video compositors, audio mastering suites, and content layout tools, must adopt standardized metadata retention rules to ensure provenance continuity across the creative pipeline [36]. This requires vendor collaboration to ensure that metadata encoding is preserved during transformations such as transcoding, trimming, resolution scaling, and file format conversion [39].

Table 2 provides a matrix of integration requirements for major generative AI vendors, comparing support for metadata embedding, cryptographic signing, dataset documentation disclosure, and API-based audit interoperability.

Cloud platforms and content delivery networks must additionally support automated verification of provenance upon upload or distribution, preventing the circulation of assets with missing or corrupted metadata [38]. Achieving compatibility across platforms ensures that provenance tracking is not dependent on isolated technical ecosystems, but operates consistently across the full media lifecycle from creation to publication, licensing, archival storage, and dispute adjudication [40].

Table 2. Integration Requirements for Major Generative AI Vendors

Vendor / Platform	Support for Metadata Embedding at Generation	Availability of Cryptographic Signing or Tamper-Evident Markers	Disclosure of Training Dataset Documentation (Scope, Source, Licensing)	API-Based Audit and Provenance Verification Interoperability
OpenAI (e.g., GPT, DALL·E)	Partial – provides limited origin indicators; metadata not embedded across export formats	Limited – watermarking features experimental and non-universal	Limited – high-level dataset descriptions provided; detailed training lineage not disclosed	Moderate – API allows programmatic logging but lacks standardized provenance exchange formats
Google (e.g., Imagen, Gemini / Vertex AI)	Partial – metadata tagging available in some controlled enterprise deployments; absent in consumer tools	Limited – integrity features exist but not configured for universal audit provenance	Limited – dataset sourcing explained generally; granular traceability remains unavailable	Moderate – enterprise APIs permit model usage auditing, but no cross-platform provenance standard
Adobe Firefly / Adobe Creative Cloud Ecosystem	Strong – supports embedded provenance tagging (Content Credentials) across export pathways within ecosystem	Moderate-High – digitally signed credentials supported when retained in Adobe-native workflows	Limited-Moderate – indicates when synthetic elements are present; training dataset specifics selective	Low-Moderate – interoperability decreases when content leaves Adobe suite or is transcoded externally
Midjourney	Very Weak – no embedded metadata included in generated outputs	None – no cryptographic or watermark-based authenticity controls	Very Limited – dataset sources not transparently disclosed; no traceable lineage provided	Very Low – no audit API or provenance verification interface
StabilityAI (e.g., Stable Diffusion)	Variable – depends on front-end interface; open-source forks often omit metadata entirely	None by default – requires external or custom tooling for signature binding	Very Limited – dataset transparency discussed broadly; lacks authoritative dataset registry identifiers	High Potential – open architecture allows third-party provenance layers, but integration is not standardized
Canva, Runway	Moderate – some	Low – watermarkin	Limited – high-level	Moderate – workflow

International Journal of Computer Applications Technology and Research
Volume 11–Issue 12, 711 - 723, 2022, ISSN:-2319–8656
DOI:10.7753/IJCATR1112.1031

Vendor / Platform	Support for Metadata Embedding at Generation	Availability of Cryptographic Signing or Tamper-Evident Markers	Disclosure of Training Dataset Documentation (Scope, Source, Licensing)	API-Based Audit and Provenance Verification Interoperability
ML, and Integrated Creative Tool Platforms	export formats retain authoring metadata; synthetic origin often not flagged	g optional and not tied to tamper-evident verification	training descriptions shared; no granular dataset lineage	APIs available, but provenance requires custom implementation

7. LEGAL, ECONOMIC, AND ETHICAL IMPLICATIONS

7.1 Impacts on Ownership, Licensing, and Creative Labor Markets

The introduction of standardized provenance metadata reshapes ownership claims and licensing structures by making the roles of human and algorithmic contributors explicitly visible [41]. Traditionally, ownership in creative industries has been based on demonstrable authorship, enabling individuals or institutions to assert exclusive rights over reproduction and distribution. However, when generative AI systems participate in content production, ownership becomes distributed among model developers, dataset curators, platform operators, and end-users providing prompts or editing inputs [38]. Provenance standards clarify these relationships by recording the specific contributions of each actor, thereby influencing how licensing fees, royalties, and usage rights are allocated [43].

For creative labor markets, provenance plays an equally consequential role. Creative workers increasingly operate alongside automated systems that accelerate production and reduce the labor intensity of output creation [40]. Without provenance, the value of human creative input may become obscured, making it difficult for workers to demonstrate authorship, negotiate credit, or secure fair compensation [39]. Provenance metadata re-establishes visibility by providing granular attribution records that identify human creative decisions, revisions, and evaluative judgments within hybrid workflows [44]. This transparency supports contractual negotiations, professional accreditation, and collective bargaining efforts by tying compensation to documented contribution rather than assumptions about authorship. Moreover, provenance structures can sustain diversity of creative expression by preventing the market from conflating synthetic replication with original artistic labor [45]. In this sense, metadata standards do not merely enforce compliance; they help preserve the economic viability and cultural legitimacy of creative professions.

7.2 Enforcement Mechanisms and Judicial Evidentiary Standards

Provenance metadata strengthens enforcement by providing verifiable, machine-readable evidence of authorship, ownership, and transformation histories that meet judicial evidentiary standards [40]. Courts require reliable documentation to resolve disputes regarding originality, infringement, licensing violations, and derivative works. Without structured provenance, litigants may rely on subjective testimony or unverifiable claims about creative process and intent [42]. By contrast, metadata-secured provenance records offer chronological and cryptographic proof of content generation, modification, and distribution, reducing ambiguity in legal interpretation [44].

Enforcement mechanisms must integrate automated checks that verify metadata integrity, detect tampering, and flag discrepancies between declared and actual authorship conditions [38]. These mechanisms are essential for regulatory agencies adjudicating claims, licensing bureaus managing copyright registrations, and professional certification bodies verifying creative contributions. Additionally, provenance allows legal authorities to distinguish between deliberate infringement and incidental similarity that may arise from the statistical generalization characteristics of generative models [41].

Judicial standards also require that provenance records maintain continuity across platforms and jurisdictions, ensuring that evidence remains admissible regardless of how digital assets circulate [45]. Metadata standards support this by defining uniform formats for audit logs, contributor identifiers, and transformation histories. Furthermore, institutional support for provenance auditing reduces reliance on proprietary platform transparency, preventing conflicts of interest where platform operators may benefit from ambiguous authorship conditions [39]. Ultimately, provenance-aligned enforcement strengthens the legitimacy and reliability of intellectual property adjudication in environments where human and algorithmic creativity intersect.

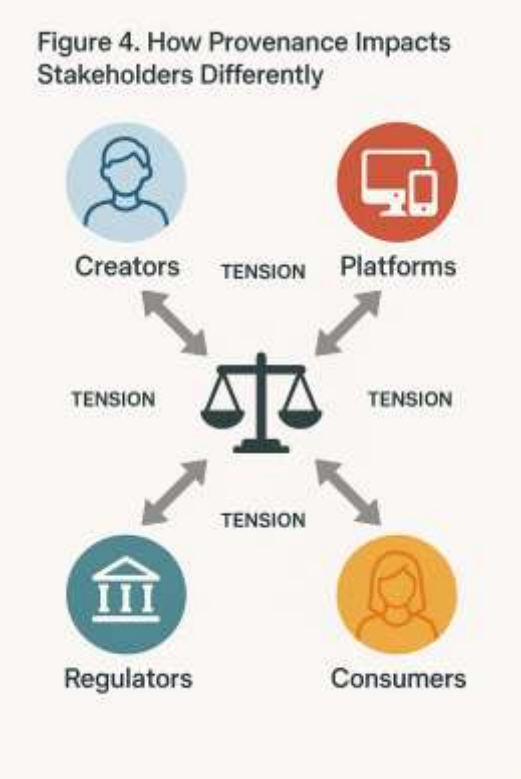
7.3 Balancing Transparency, Privacy, and Innovation Incentives

Standardized provenance frameworks must balance transparency with privacy and innovation concerns [38]. While granular attribution benefits creators and regulators, extensive disclosure of prompts, datasets, and model configurations may reveal proprietary workflows or personal creative patterns that individuals or organizations consider sensitive [44]. Therefore, access to provenance fields should follow a tiered permissions model distinguishing between public disclosure, rights-holder review, and regulatory audit access [43].

Additionally, preserving innovation incentives requires ensuring that provenance mechanisms do not impose excessive compliance burdens or inhibit experimentation in creative workflows [40]. Efficient automation and interoperable standards reduce overhead while maintaining integrity.

Figure 4 depicts how provenance impacts stakeholders differently, highlighting potential tensions and points of negotiated balance among creators, platforms, regulators, and consumers.

By designing provenance infrastructures that respect confidentiality while supporting accountability, systems can promote trust without constraining creative growth [45].



8. CONCLUSION AND FUTURE DIRECTIONS

8.1 Summary of Contributions

This article has examined the foundational need for standardized metadata protocols to ensure transparent provenance tracking for AI-generated media across federal intellectual property and regulatory systems. It established how the rise of generative AI has disrupted traditional authorship assumptions, creating uncertainty in ownership, creative attribution, and legal recognition of contribution. The discussion traced how existing IP protections, metadata practices, and evidentiary procedures are insufficient for environments where media may be produced, transformed, or distributed through automated and multi-party workflows. The proposed national framework emphasized architectural principles, field specifications, tamper-evident encoding techniques, interoperability requirements, and governance mechanisms necessary to restore credibility and traceability across creative ecosystems. The article also outlined implementation pathways spanning copyright offices, patent evaluation systems, archival infrastructures, commercial platforms, and media production environments. Finally, it considered the broader implications for labor markets, licensing systems, enforcement procedures, and ethical balancing of transparency and privacy. Collectively, these contributions demonstrate that provenance is not merely a technical detail but a structural enabler of fairness, accountability, and sustainability in AI-mediated cultural production.

8.2 Importance of Nationwide Alignment and Interoperability

Nationwide alignment is essential to ensure that provenance metadata functions consistently across institutions, platforms, and sectors. Without interoperability, provenance records risk fragmentation, producing inconsistencies that weaken legal credibility and disrupt cross-agency coordination. A unified national standard prevents the emergence of incompatible vendor formats, proprietary disclosure systems, and parallel regulatory interpretations. It enables content to move seamlessly from creation to licensing, publication, archival storage, and judicial review without loss of attribution integrity. Interoperability also promotes fairness by ensuring that creative workers, regardless of the tools or platforms they use, are recognized and protected under the same evidentiary frameworks. For businesses and cultural institutions, interoperability reduces compliance overhead by establishing shared expectations for metadata retention, audit mechanisms, and authentication procedures. For federal agencies, alignment reduces burdens on administrative review processes and strengthens the evidentiary chain of custody. At a societal level, nationwide interoperability supports public trust by enabling transparent verification of media authenticity, especially in environments increasingly affected by misinformation, synthetic media, and automated content streams. Thus, alignment is not only technically advantageous it is foundational to maintaining intellectual, cultural, and civic integrity in digital communication environments.

8.3 Roadmap for Ongoing Research, Standards Development, and Policy Action

Future research should focus on refining metadata field vocabularies, developing scalable cryptographic provenance-binding methods, and improving automated audit tools capable of detecting undisclosed AI involvement. Cross-disciplinary collaboration among computer scientists, legal scholars, archivists, creative labor representatives, and federal agencies will be necessary to ensure that standards reflect both technological feasibility and cultural legitimacy. Pilot programs can be established in partnership with federal copyright registries, academic digital archives, and large-scale creative platforms to test metadata persistence, interpretability, and reliability in real production workflows.

Policy development should prioritize regulatory clarity regarding thresholds of human involvement required for authorship recognition, obligations for platform-level provenance disclosure, and procedures for adjudicating disputes involving hybrid AI-human creative artifacts. Long-term governance should include periodic review committees tasked with updating standards to reflect evolving model architectures, ethical considerations, and industry practices. Public education initiatives will also be necessary to ensure that creators, users, and institutions understand how provenance systems function and how they support rights protection. A coordinated roadmap combining research, standard-setting, and policy action will enable provenance infrastructures to develop in ways that are sustainable, equitable, and responsive to the ongoing transformation of digital creativity.

9. REFERENCE

1. Jahankhani H, Kendzierskyj S, Montasari R, Chelvachandran N, editors. Social Media Analytics, Strategies and Governance. Taylor & Francis Group; 2022 Aug 18.
2. Hewage CT, Khattak SK, Ahmad A, Mallikarachchi T, Ukwandu E, Bentotahewa V. Multimedia privacy and security landscape in the wake of ai/ml. Social Media Analytics, Strategies and Governance. 2022 Aug 18:203-28.
3. Benhamou Y, Ferland J. Digitization of GLAM collections and copyright: Policy paper. GRUR International. 2022 May 1;71(5):403-21.
4. Benhamou Y, Ferland J, Renold MA. Digitization of GLAM Collections: Policy Paper. Available at SSRN 3963359. 2021 Nov 14.
5. Sharp AJ. Head in the Bitcloud: A Discussion on the Copyrightability and Ownership Rights in Generative Digital Art and Non-Fungible Tokens. San Diego L. Rev.. 2022;59:637.
6. Emmanuel Damilola Atanda. EXAMINING HOW ILLIQUIDITY PREMIUM IN PRIVATE CREDIT COMPENSATES ABSENCE OF MARK-TO-MARKET OPPORTUNITIES UNDER NEUTRAL INTEREST RATE ENVIRONMENTS. International Journal Of Engineering Technology Research & Management (IJETRM). 2018Dec21;02(12):151–64.
7. Beatty JS. Technologies of Convenience. Algorithmic Culture: How Big Data and Artificial Intelligence Are Transforming Everyday Life. 2020 Nov 24:141.
8. Naylor M. Insurance transformed. Cham: palgrave macmillan; 2017.
9. Bonadio E, Lucchi N. How Far Can Copyright Be Stretched?-Framing the Debate on Whether New and Different Forms of Creativity Can Be Protected. Intellectual Property Quarterly. 2019 Apr 1;2019(2):115-35.
10. Nixon L, Dasiopoulou S, Evain JP, Hyvönen E, Kompatsiaris I, Troncy R. Multimedia, broadcasting, and eculture. InHandbook of Semantic Web Technologies 2011 (pp. 911-975). Springer, Berlin, Heidelberg.
11. Van de Sompel H, Klein M. Introducing the memento tracer framework for scalable high-quality web archiving. IniPRES 2019 2019.
12. Rajecki K. Repository and Preservation Storage Architecture. IniPRES 2008 2008.
13. Lee C, Woods K. Diverse digital collections meet diverse uses: applying natural language processing to born-digital primary sources. IniPRES 2017 2017.
14. Woods K, Chassanoff A, Lee CA. Managing and transforming digital forensics metadata for digital collections. IniPRES 2013 2013.
15. Gheran BF, Villarreal-Narvaez S, Vatavu RD, Vanderdonckt J. RepliGES and GEStory: visual tools for systematizing and consolidating knowledge on user-defined gestures. InProceedings of the 2022 International Conference on Advanced Visual Interfaces 2022 Jun 6 (pp. 1-9).
16. Rumbidzai Derera. HOW FORENSIC ACCOUNTING TECHNIQUES CAN DETECT EARNINGS MANIPULATION TO PREVENT MISPRICED CREDIT DEFAULT SWAPS AND BOND UNDERWRITING FAILURES. International Journal of Engineering Technology Research & Management (IJETRM). 2017Dec21;01(12):112–27.
17. Bakos A, Miksa T, Rauber A. Research data preservation using process engines and machine-actionable data management plans. InInternational Conference on Theory and Practice of Digital Libraries 2018 Sep 5 (pp. 69-80). Cham: Springer International Publishing.
18. Burny N, Vanderdonckt J. GUIMETRICS: An Extensible Cloud-based Application for Automatic Computation of GUI Visual Design Measures. InICSOFT 2021 (pp. 505-512).
19. Shrestha A, Mahmood A. Review of deep learning algorithms and architectures. IEEE access. 2019 Apr 22;7:53040-65.
20. Pröll S, Rauber A. Enabling reproducibility for small and large scale research data sets. D-Lib Magazine. 2017 Jan;23(1/2).
21. Daniel ONI. TOURISM INNOVATION IN THE U.S. THRIVES THROUGH GOVERNMENTBACKED HOSPITALITY PROGRAMS EMPHASIZING CULTURAL PRESERVATION, ECONOMIC GROWTH, AND INCLUSIVITY. International Journal Of Engineering Technology Research & Management (IJETRM). 2022Dec21;06(12):132–45.
22. Atanda ED. Dynamic risk-return interactions between crypto assets and traditional portfolios: testing regime-switching volatility models, contagion, and hedging effectiveness. *International Journal of Computer Applications Technology and Research*. 2016;5(12):797–807.
23. Ibitoye J, Fatanmi E. Self-healing networks using AI-driven root cause analysis for cyber recovery. International Journal of Engineering and Technical Research. 2022 Dec;6: doi:10.5281/zenodo.16793124.
24. Takuro KO. Assessing the legal and regulatory implications of blockchain technology on smart

contracts, digital identity, and cross-border transactions. *World Journal of Advanced Research and Reviews*. 2022;16(3):1426-1442. doi:10.30574/wjarr.2022.16.3.1350.

25. Owens T, Sands AE, Reynolds E, Neal J, Mayeaux S. Guest Editorial Libraries Advancing the National Digital Platform. D-Lib Magazine. 2017 May;23(5/6).
26. Ibitoye JS. Securing smart grid and critical infrastructure through AI-enhanced cloud networking. International Journal of Computer Applications Technology and Research. 2018;7(12):517-529. doi:10.7753/IJCATR0712.1012.
27. Derera R. Machine learning-driven credit risk models versus traditional ratio analysis in predicting covenant breaches across private loan portfolios. *International Journal of Computer Applications Technology and Research*. 2016;5(12):808-820. doi:10.7753/IJCATR0512.1010.
28. Yu F, Xiu X, Li Y. A survey on deep transfer learning and beyond. Mathematics. 2022 Oct 3;10(19):3619.
29. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S. Language models are few-shot learners. Advances in neural information processing systems. 2020;33:1877-901.
30. Jadon S, Jadon A. An overview of deep learning architectures in few-shot learning domain. arXiv preprint arXiv:2008.06365. 2020 Aug 12.
31. Zhu X, Gu Y, Xiao Z. Herbkg: Constructing a herbal-molecular medicine knowledge graph using a two-stage framework based on deep transfer learning. Frontiers in Genetics. 2022 Apr 27;13:799349.
32. Li J, Sun A, Han J, Li C. A survey on deep learning for named entity recognition. IEEE transactions on knowledge and data engineering. 2020 Mar 17;34(1):50-70.
33. Deng C, Ji X, Rainey C, Zhang J, Lu W. Integrating machine learning with human knowledge. Iscience. 2020 Nov 20;23(11).
34. De Lange M, Aljundi R, Masana M, Parisot S, Jia X, Leonardis A, Slabaugh G, Tuytelaars T. A continual learning survey: Defying forgetting in classification tasks. IEEE transactions on pattern analysis and machine intelligence. 2021 Feb 5;44(7):3366-85.
35. Izacard G, Lewis P, Lomeli M, Hosseini L, Petroni F, Schick T, Dwivedi-Yu J, Joulin A, Riedel S, Grave E. Few-shot learning with retrieval augmented language models. arXiv preprint arXiv:2208.03299. 2022 Aug;1(2):4.
36. Siebers P, Janiesch C, Zschech P. A survey of text representation methods and their genealogy. IEEE Access. 2022 Sep 12;10:96492-513.
37. Barlaug N, Gulla JA. Neural networks for entity matching: A survey. ACM Transactions on Knowledge Discovery from Data (TKDD). 2021 Apr 21;15(3):1-37.
38. Khatun A, Rahman A, Islam MS, Chowdhury HA, Tasnim A. Authorship attribution in bangla literature (aabl) via transfer learning using ulmfit. Transactions on Asian and Low-Resource Language Information Processing. 2020.
39. Wörmann J, Bogdoll D, Brunner C, Bührle E, Chen H, Chuo EF, Cvejoski K, van Elst L, Gottschall P, Griesche S, Hellert C. Knowledge augmented machine learning with applications in autonomous driving: A survey. arXiv preprint arXiv:2205.04712. 2022 May 10.
40. Adadi A. A survey on data-efficient algorithms in big data era. Journal of Big Data. 2021 Jan 26;8(1):24.
41. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow PM, Zietz M, Hoffman MM, Xie W. Opportunities and obstacles for deep learning in biology and medicine. Journal of the royal society interface. 2018 Apr 30;15(141):20170387.
42. Wankhade M, Rao AC, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. Artificial Intelligence Review. 2022 Oct;55(7):5731-80.
43. Joshi M, Pal A, Sankarasubbu M. Federated learning for healthcare domain-pipeline, applications and challenges. ACM Transactions on Computing for Healthcare. 2022 Nov 3;3(4):1-36.
44. Ye E, Bai X, O'Hare N, Asgarieh E, Thadani K, Perez-Sorrosal F, Adiga S. Multilingual taxonomic web page classification for contextual targeting at yahoo. InProceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining 2022 Aug 14 (pp. 4372-4380).
45. Volk MJ, Lourentzou I, Mishra S, Vo LT, Zhai C, Zhao H. Biosystems design by machine learning. ACS synthetic biology. 2020 Jun 2;9(7):1514-33.