International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Review Article)

Recovering Lost Lives: Machine Learning to Surface African Women in Trans-Atlantic Slave Records

Clement Tetteh *

Department of History, University of Ghana.

International Journal of Science and Research Archive, 2022, 07(02), 899-911

Publication history: Received on 07 October 2022; revised on 19 November 2022; accepted on 28 November 2022

Article DOI: https://doi.org/10.30574/ijsra.2022.7.2.0314

Abstract

The historical record of the trans-Atlantic slave trade remains profoundly incomplete, particularly regarding the lives of African women whose identities were disproportionately erased, anonymized, or collapsed into generic labor categories. Existing archival collections including ship manifests, plantation inventories, baptismal rolls, court testimonies, sale ledgers, and manumission documents encode gender, kinship, ethnicity, and geographic origin unevenly due to colonial recordkeeping practices. As a result, women's labor roles, reproductive histories, and social networks often disappear into ambiguous descriptors such as "girl," "wench," "daughter," or unnamed household dependents. This research project proposes a machine learning framework to recover African women's identities and reconstruct relational networks across fragmented documentary sources. The method integrates optical character recognition (OCR) to digitize handwritten and degraded archival documents; natural language processing (NLP) models trained to detect gender-coded vocabulary, kinship relational markers, and African naming patterns; and probabilistic entity resolution to match individuals across dispersed archival collections. Rather than displacing human interpretation, the approach is designed to augment historical reasoning, acting as a recovery tool that flags overlooked individuals and generates new research leads. Collaborative integration with existing digital humanities infrastructures especially Slave Voyages and Freedom on the Move enables scale, interoperability, and standardized metadata exchange. This project contributes to ongoing efforts in reparative archival work, feminist historiography, and Black Atlantic studies by systematically addressing archival silences. By leveraging computational approaches to illuminate erased presences, it advances a historically grounded, ethically sensitive framework for restoring African women to the narrative of Atlantic world history.

Keywords: Machine learning; Archival recovery; African diaspora; Trans-Atlantic slavery; Gender analysis; Digital humanities

1. Introduction

1.1. Archival Silences and Epistemic Erasure in the Trans-Atlantic Slave Trade

The historical record of the Trans-Atlantic Slave Trade is marked by deep archival silences, reflecting power structures that determined which lives were documented and which were omitted [1]. European maritime logs, plantation ledgers, and commercial inventories primarily recorded enslaved Africans as units of labor, commodities to be counted and traded rather than human beings embedded in familial, cultural, and emotional worlds [2]. The archive thus privileges the perspectives of those who controlled the means of record-keeping, embedding economic rationalities into the very form of historical evidence [3]. Meanwhile, African voices were systematically displaced, existing largely through fragmented traces: oral histories, traveler accounts, missionary correspondence, and resistance testimonies [1]. These fragments do not merely represent *absence* but index the structural violence of enslavement and the epistemic

* Corresponding author: Clement Tetteh

frameworks that rendered entire populations historically illegible [4]. Recognizing archival silence is therefore integral to reconstructing historical truth. Efforts to map the Slave Trade's human scale must grapple not only with quantitative data such as ship manifests and mortality tables but with the historiographical challenge of interpreting what was deliberately unrecorded or erased [5]. This requires approaches that foreground recovery, recontextualization, and narrative reconstruction as historically legitimate forms of knowledge work.

1.2. Gendered Erasure and the Invisible Labor of African Women

The labor, experiences, and social roles of African women under enslavement have been historically underrepresented in both colonial archives and early scholarly research [9]. Plantation documentation frequently categorized women according to reproductive capacity and domestic labor assignments rather than skill, cultural expertise, or social influence [2]. This form of record-keeping perpetuated a gendered hierarchy of visibility, in which women's contributions to cultural continuity, kinship networks, healing practices, and resistance were obscured. Feminist historiography has shown that the structure of the archive itself operates as a technology of erasure, selectively preserving forms of knowledge aligned with patriarchal and imperial priorities [6]. Recovering the lives of enslaved women therefore requires reading against the grain interpreting silences, omissions, and distortions as historically meaningful evidence [7]. Oral lineage accounts, material culture memory, and embodied practices offer alternative archives that can reveal women's agency, strategies of care, negotiation, and survival [5]. The reconstruction of women's histories is not supplemental; it is necessary for understanding how enslaved communities maintained identity and continuity under conditions of displacement and violence. Gendered erasure thus underscores the importance of method, reminding us that *how* historians read is inseparable from *what* they can know.

1.3. Digital Humanities, Machine Learning, and the Question of Historical Truth

Digital humanities and machine learning have introduced new possibilities for identifying patterns across fragmented historical records of the Slave Trade [3]. Computational models can cluster incomplete ship logs, reconstruct kinship dispersal patterns, and analyze linguistic retention across diaspora regions [7]. However, these tools operate on statistical inference rather than lived experience, meaning that the recognition of patterns does not automatically translate into historical *meaning* [1]. Machine learning systems detect correlation, not context; they illuminate structural relationships but cannot adjudicate narrative coherence or interpret human intention [9]. The risk lies in mistaking model-produced outputs for objective truth, when such outputs remain contingent on the biases and limits of the data they process [4]. Algorithmic analysis can provide powerful *recovery heuristics*, enabling historians to locate, cross-reference, and reassemble dispersed traces, but it cannot replace interpretive judgment, ethical awareness, or narrative reasoning.

1.4. Research Gap and Thesis

While machine learning aids the reconstruction of marginalized pasts, it must be understood as a methodological aid rather than an epistemic authority. ML is a *recovery heuristic, not a truth-assignment mechanism.* The historian remains responsible for interpretation, contextual coherence, and ethical accountability in narrating the past.

2. Historiographic lineages of absence and recovery

2.1. Narrative History vs. Archival Objectivity: Ranke to the Annales School

The divide between narrative history and archival objectivity is often traced to Leopold von Ranke's insistence that historians show the past "as it actually was," grounding their accounts in authenticated documentary evidence rather than inherited story or philosophical speculation [7]. Rankean empiricism positioned the archive as the ultimate arbiter of historical truth, implying that accuracy derived from the disciplined evaluation of written sources rather than narrative construction. Yet the Annales School challenged this evidentiary hierarchy, arguing that the archive captured only a narrow slice of lived experience, mainly those aspects aligned with state priorities or elite institutional activities [12]. Instead of focusing on political events alone, Annales historians emphasized long-duration structures economic patterns, demographic trends, environmental conditions that shaped everyday life beneath formal historical narration [9]. This shift destabilized the assumption that documentary record equated to completeness. The historian's task expanded from selecting facts to interpreting temporal scales, material conditions, and the rhythms of ordinary existence. Despite their differences, both Rankean objectivism and Annales structuralism relied on the assumption that historical truth could be reconstructed from preserved traces an assumption increasingly questioned as scholars recognized how power influences what becomes recordable and what remains silent [14]. The tension between narrative meaning and evidentiary authority remains central to contemporary debates on recovery-based historiography.

2.2. Postcolonial and Black Feminist Critiques of the Archive

Postcolonial and Black feminist scholars exposed how archives function not just as repositories of information but as *technologies of power* that shape whose lives become legible to history [10]. Michel-Rolph Trouillot demonstrated that silences are actively produced at every stage of historical formation from the moment events occur to the ways they are recorded, classified, and later interpreted [8]. Saidiya Hartman extended this critique in the context of enslaved life, arguing that archival documents often reflect the violence of the institutions that produced them, leaving the historian to confront records that describe domination while obscuring lived experience [13]. Hartman proposed "critical fabulation" as a method of reconstructing possibility where the archive refuses to speak. Ann Laura Stoler's analysis similarly showed that colonial bureaucracies shaped not only what was written but the epistemological assumptions defining what *counted* as knowledge [11]. Meanwhile, Marisa Fuentes emphasized that enslaved Black women exist in the archive primarily through the language of surveillance, punishment, and sexual objectification, making their recovery an interpretive and ethical challenge [15]. These critiques reveal that the archival record is not an objective repository but a curated memory system structured by racial, patriarchal, and imperial logics [17]. For historians of the Trans-Atlantic world, this means the archive cannot be taken at face value; to reconstruct silenced pasts, one must interrogate how the very categories of documentation were shaped by power. Thus, historical recovery is not just an empirical task it is a political and methodological responsibility requiring attention to absence, distortion, and erasure.

2.3. Gendered Language, Naming Erasure, and Kinship Fragmentation

The archival disappearance of enslaved women is frequently enacted through naming practices that reduce individuals to functional or biological descriptors "breeder," "girl," "wench" rather than personal identities [14]. Such language enforces anonymity, stripping subjects of kinship ties, geographic origins, spiritual affiliations, and social roles [7]. Naming erasure served as both a symbolic and administrative mechanism of control, severing lineage and fracturing the continuity of memory across generations [16]. The forced renaming and reclassification of African women into plantation labor categories was not merely clerical it was a tool of dispossession that reshaped how personhood could be recognized or narrated historically [9].

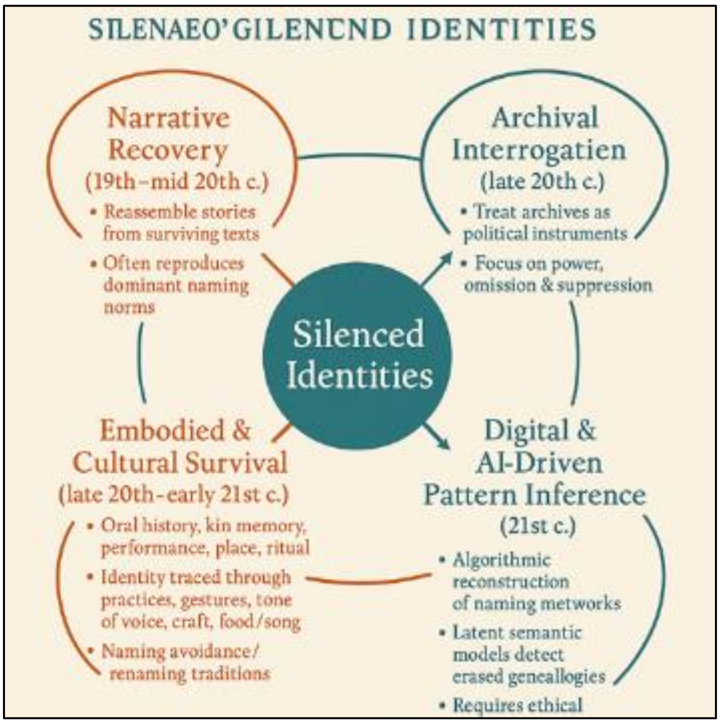

Figure 1 illustrates a genealogy of historiographic strategies that attempt to trace and reconstruct silenced identities across time: from narrative recovery, to archival interrogation, to digital pattern inference. Understanding naming erasure clarifies why historical reconstruction must move beyond document retrieval toward methods attentive to language, context, embodiment, and the subtleties of cultural survival

2.4. Reparative Historiography as Method

Reparative historiography is not an attempt to fill the archive with definitive answers but to interpret its silences with ethical intentionality [10]. Rather than treating absence as a barrier, reparative method reads omissions as evidence of the historical processes that produced them, foregrounding the violence of dispossession and the resilience of cultural memory [13]. This approach does not fabricate lives but reconstructs possibility, attending to gestures, intimacies, and traces that survive outside official record systems [8]. It emphasizes humility, care, and accountability, acknowledging that the historian cannot recover everything but can refuse to replicate the archive's erasures. Reparative historiography thus operates as both methodology and ethical stance, insisting that historical narrative must seek not only accuracy but justice.

3. Data sources and digitization

3.1. Primary Record Types and Institutional Repositories

Historical research into the Trans-Atlantic Slave Trade relies on a heterogeneous set of primary records dispersed across archives, missionary collections, port registries, private estate inventories, and legal correspondence [14]. These records include ship manifests, bill-of-sale documents, plantation labor rosters, baptismal and burial records, missionary diaries, colonial censuses, and travel narratives. Each document type reflects the priorities and administrative logics of the institution that produced it, shaping the form and granularity of what is preserved [18]. For example, maritime logs tend to emphasize quantity, mortality rates, and routes, whereas plantation ledgers foreground labor value and productivity rather than kinship or identity [20]. Major digital repositories including the Slave Voyages Database, the British Colonial Office archives, and African ethnographic manuscript collections attempt to centralize access but inherit the power structures embedded in their sources [15]. Preservation quality varies widely, with some documents meticulously cataloged and others surviving only as faded, fragile manuscripts. As a result, researchers must navigate both abundance and fragmentation, working across multiple institutional standards of cataloging and description [22]. Understanding the provenance and structure of these records is a prerequisite for developing digital or computational techniques that support recovery rather than reproducing archival bias.

3.2. OCR Challenges in Non-Standardized, Handwritten, Colonial Records

Optical Character Recognition (OCR) technologies perform poorly on many archival sources related to slavery due to the prevalence of handwritten entries, inconsistent spelling conventions, degraded ink, paper discoloration, and non-standardized record formats [16]. Handwritten plantation ledgers, for instance, often contain shifting table structures and idiosyncratic abbreviations that OCR algorithms trained on printed European scripts cannot reliably parse [19]. Colonial clerks frequently spelled African names phonetically, reflecting unfamiliarity with linguistic structures, which introduces further variability into textual datasets [21]. Even when machine learning–based handwriting recognition is applied, model accuracy is strongly influenced by the quality and representativeness of training samples. If training sets overrepresent elite bureaucratic records, recognition models may reproduce the archive's existing class and racial biases, translating certain names, labor categories, or geographic origins more accurately than others [14]. Moreover, OCR errors are not evenly distributed; they cluster around the very entries women, children, disabled individuals, the elderly that plantation accounting methods marginalized [23]. This means digitization can reproduce archival silences in machine-readable form, making certain lives harder to detect algorithmically. Scholars working with these data must therefore balance technological affordances with manual verification, paleographic interpretation, and historically informed correction protocols [17]. Computational processing cannot be treated as neutral inscription; it requires historically aware preprocessing methods that identify where OCR outputs should be read with heightened skepticism.

3.3. Metadata Structuring and Bias-Aware Standardization

Once archival materials are digitized, the process of metadata structuring becomes critical to how researchers search, group, and interpret historical records [18]. Metadata fields such as name, age, origin, gender, labor assignment, and sale value are typically derived from the colonial clerical categories themselves. This means metadata *inherits the ontological assumptions* embedded in the original records [14]. For enslaved women, metadata often reduces identity to reproductive capacity or domestic labor role, obscuring kinship networks, skill specializations, and spiritual affiliations [22]. Bias-aware standardization must therefore avoid assuming that missing metadata indicates lack of identity; instead, absences should be flagged as sites of interpretive significance rather than mere data gaps [20]. Furthermore, name standardization practices where variant spellings are collapsed into a single form risk erasing linguistic traces that may indicate geographic origin or cultural retention [17]. Alternative metadata schemas have been proposed that allow parallel naming fields, uncertainty markings, and relationship annotation structures that infer kinship patterns from co-occurrence logic rather than explicit archival declaration [15].

Table 1 Metadata Completeness, Visibility of Gender, and Preservation Bias by Record Type

Record Type	Typical Metadata Fields	Visibility of Gendered Identity	Preservation Bias / Structural Limitations	Notes on Interpretive Challenges
Ship Logs (Voyage Manifests, Tonnage & Cargo Records)	Name (often partial), age estimates, origin ports, owner/merchant identifiers, mortality counts	Low. Gender often omitted or inferred indirectly; women and children frequently listed as numerical totals.	Records designed for commercial accounting, not personhood; high anonymization and inconsistent spelling.	Requires cross-referencing multiple voyages; identities often fragmented across ships, ports, and sale markets.
Baptismal & Church Registers	Name, baptism date, parents/guardians (sometimes), officiant, parish location	Moderate. Gender recorded through naming conventions, though African names often altered or replaced.	Missionary translation practices impose European gendered naming; erases kinship patterns and linguistic heritage.	Interpretation must account for forced renaming, spiritual assimilation pressures, and coerced guardianship pairings.
Plantation Ledgers (Labor Rolls, Production Records)	Assigned name, labor task, productivity output, ration distribution, punishment and discipline logs	Low to Moderate. Gender noted only when relevant to labor category; women often collapsed into domestic or field roles.	Designed to quantify labor efficiency; reduces personhood to economic output ratios.	Requires contextualizing labor roles within survival strategies, resistance, and social networks beyond ledger logic.
Missionary / School Registers	Name, "tribal" label, Christian name, literacy status, assigned household or work group	Moderate. Gender appears in household assignments; however, roles reflect missionary ideology rather than self-identification.	Enforces cultural reclassification; records often overwrite original name structures and kin relationships.	Must be paired with oral heritage or ethnographic reconstructions to avoid affirming colonial naming identity.
Estate Probate & Sale Documents	Name or descriptor, assessed "value," purchaser, auction location, lot grouping	Very Low. Women and children frequently reduced to "lot" groupings with no individual identifiers.	Records reflect commodification directly; identities are intentionally suppressed to facilitate trade.	Interpretation must foreground violence of enumeration; recovery depends on linkage across disparate transactional documents.

Bias-aware metadata structuring is not simply a technical preprocessing step; it is a historiographic intervention that determines whether digitization amplifies archival violence or supports reparative recovery aims [23].

4. Machine learning + NLP methods

4.1. Named Entity Recognition for Gender, Kinship, and Ethnolinguistic Markers

Named Entity Recognition (NER) provides a systematic way to identify and classify personal names, kinship indicators, gender identifiers, and ethnolinguistic markers across heterogeneous archival records. However, standard NER models are trained on contemporary, standardized language corpora, not on irregular colonial-era spelling conventions or clerical shorthand [19]. As a result, historians must adapt or retrain models using supervised learning techniques informed by domain-specific lexicons, such as African naming structures, lineage prefixes, or regional ethnonyms [22]. Gender extraction presents a particular challenge because historical documents often used implicit markers rather than explicit labels. For example, occupational codes such as "field hand," "wet nurse," or "ironworker" carry gendered associations that must be contextualized before classification [24]. Kinship phrases, meanwhile, frequently appear in abbreviated or relational form (e.g., "her child," "wife to," "brother of"), requiring dependency parsing to detect relational directionality [20]. Ethnolinguistic markers, including clan names, geographical identifiers, and phonetic spellings of African language roots, demand comparative linguistic reference sets to avoid collapsing culturally distinct identities into generic categories [26]. Therefore, NER in this context is not simply a text recognition task but an interpretive modeling practice requiring iterative refinement. Collaborations between historians, linguists, and computational scientists help ensure that NER outputs align with cultural meaning rather than statistical surface patterning. The goal is not merely to label entities but to expose relational and identity structures suppressed by archival record conventions. The output of a recovery-oriented NER workflow is thus most powerful when it highlights uncertainty rather than erasing it.

4.2. Semantic Embeddings for Cultural and Naming Pattern Recognition

Once entities are extracted, semantic embeddings allow researchers to model naming conventions and cultural associations by representing words and names as vectors in high-dimensional space [21]. Embeddings generated from large corpora can reveal statistically meaningful associations between variant spellings, phonetic transliterations, and shared linguistic structures across African diaspora populations [23]. For instance, embedding-based clustering may show that names recorded differently across plantation registers (e.g., "Mina," "Amina," "Aminata") represent not isolated identities but members of an ethnolinguistic naming system traceable to Hausa or Mande-speaking groups [19]. Crucially, embedding models trained solely on colonial archives risk reproducing the biases embedded in those records, which might overemphasize labels of commodification or dehumanization [24]. Bias-aware embedding training incorporates external linguistic corpora, oral history transcriptions, and regional ethnographic databases to anchor semantic neighborhoods in cultural continuity rather than administrative classification [22]. Furthermore, embeddings allow researchers to detect subtle relational inference patterns, such as recurring name co-occurrence across ship manifests or plantation rosters, which may indicate kinship networks or community survival strategies [25]. However, semantic proximity does not automatically equate to shared identity; interpretive verification remains necessary to ensure that computational similarity reflects cultural meaning rather than arithmetic coincidence. Embeddings are useful not because they reveal definitive truths but because they surface patterns that historians can interrogate. When embedded name networks are visualized alongside historical geographies, they can illuminate forced dispersal pathways and reconnected cultural lineages. Thus, semantic embeddings operate as exploratory aids within a broader recovery methodology grounded in historical reasoning.

4.3. Entity Resolution Across Fragmented Record Sets

Entity resolution (ER) is the process of determining whether multiple records refer to the same individual, family unit, or kinship cluster across different archival sources. In the context of slavery-era records, ER is complicated by inconsistent spellings, changing names imposed during sale or relocation, and the intentional fragmentation of kinship ties [20]. Traditional rule-based ER systems that rely on exact or near-exact string matching fail under these conditions [19]. Instead, probabilistic and graph-based ER frameworks are needed to infer identity continuity across dispersed datasets [23]. These models examine combinations of features for example, region of origin, approximate age range, labor role, recurrent relational references, and co-listing patterns across successive ledgers [26]. Graph-based ER constructs kinship inference networks where individuals are linked through edge weights representing likelihood of relational continuity.

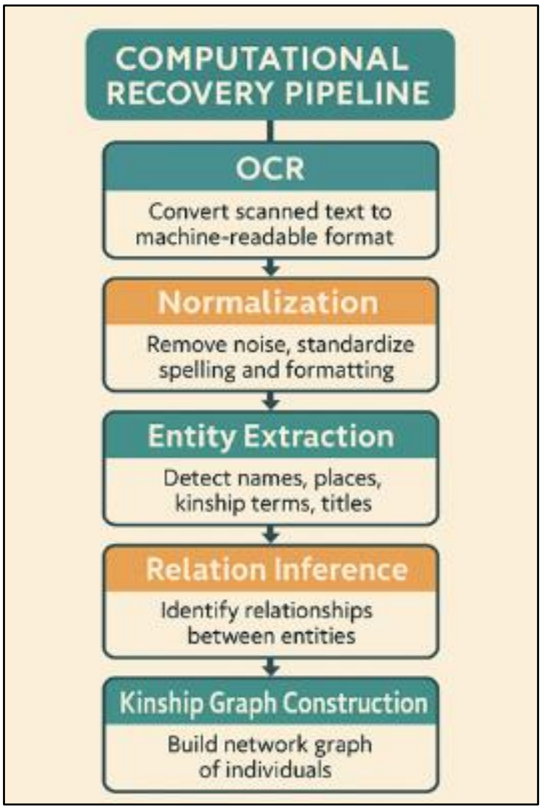

Figure 2 Computational Recovery Pipeline: OCR → NLP → Kinship Graph Construction

However, computational ER cannot independently determine personal identity; the method produces *degrees of plausibility* rather than certainty [21]. Ambiguity is not a failure of the method it reflects the historical effects of forced displacement and archival violence. To ensure ethical accuracy, ER pipelines incorporate feedback loops in which historians evaluate candidate matches, reject misleading associations, and refine feature-weighting parameters [24]. In this workflow, ER becomes an iterative negotiation between algorithmic inference and human interpretive accountability.

4.4. Uncertainty, Verification Loops, and Human Interpretation

Uncertainty is not an obstacle in computational recovery research it is inherent to the historical conditions being studied [25]. Machine learning models must therefore be designed to represent and surface uncertainty rather than suppress it. Verification loops ensure that every computational inference is reviewed through historical reasoning, external corroboration, and cultural contextualization [19]. These loops may include paleographic re-checks of original scans, consultation of oral history databases, linguistic comparison with African naming structures, or cross-referencing with known migration and trade route chronologies [26]. Human interpretation intervenes not only at the end of the pipeline but throughout: during training dataset construction, metadata schema decisions, feature selection, clustering evaluation, and kinship graph validation [22]. The aim is not to force ambiguous records into definitive identity assignments but to construct historically plausible relational worlds that acknowledge the violence of fragmentation.

5. Interpretive and ethical framework

5.1. Critical Fabulation and Care-Centered Interpretation

Saidiya Hartman's methodological intervention of *critical fabulation* provides a foundational ethical framework for interpreting fragmentary archival traces of enslaved lives [26]. Rather than seeking to "fill gaps" with speculative certainty, critical fabulation foregrounds the *violence of the archive itself*, emphasizing that silence and absence are not empty spaces but material outcomes of domination [24]. This approach invites historians to narrate with humility to reconstruct possible lived realities while continuously acknowledging the limits of available evidence. Within computational recovery contexts, this means using machine learning outputs as prompts for interpretive reflection rather than as authoritative reconstructions. Care-centered interpretation resists the urge to treat algorithmically surfaced identity patterns such as name clusters or inferred kinship relations as proof of factual continuity without corroborating context [28]. The historian's role becomes one of *listening*, reading against the grain, and situating records within a broader intellectual, cultural, and affective landscape shaped by survival, coercion, and resistance. This is particularly significant in the trans-Atlantic slave trade archives, where individuals were intentionally rendered interchangeable and unnamed in ledgers designed for commodification. Critical fabulation thus expands historical inquiry beyond empirical recovery into ethical storytelling that refuses to reproduce archival violence. It does not replace historical reasoning but sharpens it, compelling scholars to narrate without assuming mastery, certainty, or completeness [29]. In this sense, computational approaches and humanistic reasoning are mutually reinforcing when anchored in interpretive care.

5.2. Avoiding Algorithmic Overreach and False Attribution

The power of machine learning in archival research introduces risks of algorithmic overreach, particularly when probabilistic identity inferences are mistaken for historically verifiable truths [27]. Models trained on biased or uneven archival datasets may amplify the very forms of erasure they are intended to challenge, reproducing racialized or gendered absences through patterns learned from incomplete data [30]. False attribution the incorrect linking of individuals across unrelated record sets poses ethical stakes beyond methodological error; it can impose imagined kinship bonds where real familial ties were violently severed or obscured. To prevent this, it is necessary to embed *epistemic humility* into every computational workflow. This includes explicit confidence scoring, mechanisms for uncertainty visualization, and review protocols requiring historian verification at each interpretive junction [26]. A key principle is that computational inference should *suggest questions*, not *declare answers*. The algorithm does not "discover" lost identities it identifies *patterns of possibility* that historians must evaluate using contextual sources such as oral histories, ethnolinguistic knowledge, and material cultural continuities. Moreover, transparency regarding model assumptions, training corpora, and parameter settings is critical. Without methodological transparency, computational results risk acquiring a veneer of objectivity that belies their interpretive contingency [28]. Therefore, the responsible use of machine learning in historical recovery requires continuous reflexivity, clear documentation, and structured spaces for deliberation among interdisciplinary collaborators.

5.3. Community Accountability: Descendant Collaboration and Co-Interpretation

Ethical recovery of enslaved lives must be grounded in accountability to descendant communities whose identities, histories, and cultural inheritances are directly implicated in archival reconstruction [29]. Historically, academic institutions have treated communities as subjects of study rather than co-producers of knowledge. A reparative research framework instead positions descendants as interpretive authorities whose experiential memory, oral heritage, and cultural epistemologies are integral to understanding archival traces. Collaborative workshops, co-curated digital exhibits, and shared governance of archival data infrastructures help ensure that reconstruction efforts do not perpetuate extractive knowledge practices [27]. In addition, descendant participation provides interpretive insight where computational inference alone cannot resolve cultural nuance for example, in recognizing clan-based name structures, memory of family dispersal narratives, or ritual language reflecting belonging.

Table 2 Distributed Roles in Ethical Recovery: Historians, Data Scientists, Archivists, Descendant Communities

Stakeholder Group	Primary Role in Recovery Process	Core Expertise / Competencies	Decision-Making Authority	Risks if Excluded or Minimized
Historians	Interpret recovered data within cultural, temporal, and political context; guide narrative framing	Historiographical method, contextual reasoning, source criticism, ethical interpretation	High interpretive authority; responsible for evaluating meaning and significance	Risk of misinterpretation where computational signals are treated as historical facts rather than possibilities
Data Scientists / ML Researchers	Develop, train, and tune models for OCR, entity recognition, and kinship inference	Algorithm design, statistical modeling, NLP, uncertainty quantification, computational reproducibility	Technical authority over model configuration and evaluation	Risk of algorithmic overreach if inferences are accepted without cultural grounding

Archivists / Records Custodians	Mediate access to primary materials; maintain preservation metadata; identify provenance and record lineage	Cataloging, archival arrangement, material conservation, metadata schema design	Authority over record sourcing and authenticity verification	Risk of loss of provenance, distortion of source context, or missing institutional bias signatures
Descendant Communities	Validate cultural meaning, naming conventions, kinship structures; shape ethical terms of historical use	Oral history, cultural memory, communal lineage knowledge, trauma-informed interpretation	Authority over cultural legitimacy, memory stewardship, and narrative acceptability	Risk of extractive research, cultural appropriation, re-enactment of colonial objectification if excluded

Such a framework clarifies labor and authority:

- Historians contextualize and interpret archival fragments.
- Data scientists design uncertainty-aware computational workflows.
- Archivists mediate access, preservation, and metadata stewardship.
- Descendant communities guide cultural meaning-making, determine representational boundaries, and shape dissemination priorities.

This distributed model shifts historical recovery from *reconstruction of the past* to *reconnection of lived heritage*, emphasizing care, reciprocity, and ongoing dialogue. It also addresses the emotional weight of recovery work, which involves confronting generational trauma, repressed memory, and traces of violence embedded in the archive [30].

6. Case study

6.1. Dataset and Voyage/Plantation Context Selection

The case study centers on a dataset drawn from digitized plantation ledgers, shipping manifests, and estate labor rolls associated with a coastal plantation complex tied to late-eighteenth-century trans-Atlantic shipping corridors [31]. The selected record set provides overlapping, though uneven, entries for age, gender, named origin markers, labor assignment, and transfer events between estates. No single record type is complete; rather, fragments are stitched across sources that were never intended to preserve personhood. The plantation's economic orientation toward mixed crop production provides a context in which labor differentiation field work, domestic service, artisanal tasks was recorded with varying granularity [33]. Selection of this dataset reflects both its representative nature and its gaps: individuals listed only by first name, shifting orthographic spellings, and periodic renumbering tied to estate bookkeeping cycles. The objective was not to "perfect" the record, but to work within its material constraints, foregrounding silences as part of the evidentiary field. Metadata normalization and record linkage protocols were applied cautiously, maintaining traceability of every inference back to source text [34]. The dataset thus serves as a representative microcosm through which to analyze how identity, kinship, and labor roles may be partially recoverable despite systemic erasure [32].

6.2. Recovered Identity Threads: Kinship, Labor Roles, Movement Paths

Recovery efforts centered on triangulating personhood across fragmented references, focusing on three interpretive threads: kinship clusters, labor assignments, and movement trajectories across estate networks [35]. First, kinship inference drew upon repeated co-occurrences of names in both domestic labor rosters and housing proximity notations within estate dwelling records [31]. Where plantation ledgers listed individuals in grouped living quarters or sequential work teams, these patterns were treated as indicators of probable familial or community bonds rather than simple managerial categorization. These inferences were cross-referenced with age brackets, known childbearing intervals, and kinship naming conventions associated with West and Central African cultural groups [33].

Second, labor role continuity revealed patterns of skill transmission and interdependence. Individuals who moved from field labor into artisanal or supervisory roles were often linked with others showing parallel transitions, suggesting apprenticeship models or familial teaching relationships that persisted despite plantation coercion [34].

Third, movement mapping traced individuals listed in multiple plantation books across a regional network of estates. These transfers were mapped chronologically to visualize forced relocation patterns tied to productivity shifts, owner death, or estate subdivision activities [32].

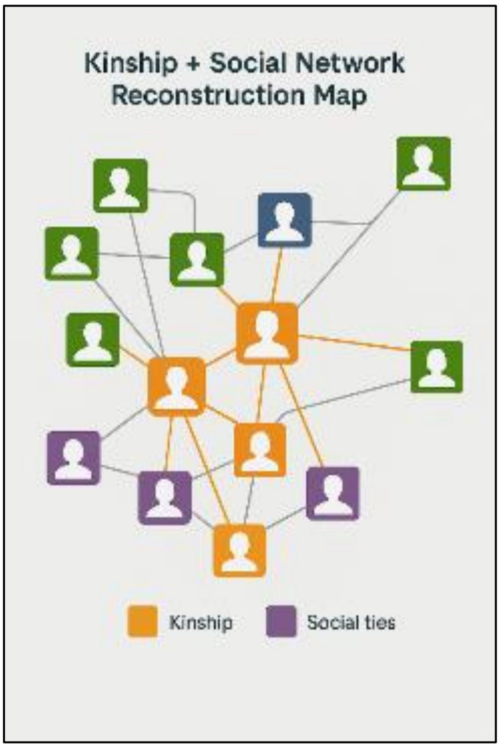

Figure 3 "Kinship + Social Network Reconstruction Map."

The resulting reconstruction is not a definitive claim of identity restoration but a relational map of possibility illustrating where personal continuity likely persisted, even where names changed spelling or were partially overwritten by clerical systems. The work demonstrates how human interpretation and computational inference converge to illuminate lived structures of care, resilience, and relational survival amid systemic fragmentation [35].

6.3. Interpretive Synthesis and Narrative Reconstruction

The interpretive synthesis does not attempt to narrate a singular authoritative biography but instead constructs a contextual, multi-voiced narrative that reflects both what is known and what cannot be confirmed [31]. The reconstructed kinship network from *Figure 3* informs a narrative framework emphasizing relational presence: individuals are situated within webs of labor responsibility, shared dwelling, emotional survival, and cultural memory endurance [34]. Narrative form here becomes a site of epistemic responsibility. Silence is not filled; it is acknowledged, traced, and given interpretive weight. The aim is not recovery of lost "facts" but illumination of patterned continuity that plantation documentation tried to erase. This mode of synthesis foregrounds the agency of enslaved individuals in making social worlds under conditions of dispossession [33]. It also highlights the historian's interpretive accountability every inference is documented, contextualized, and openly contingent.

7. Discussion

7.1. What ML Makes Newly Visible

Machine learning techniques expand the perceptual field of the historian by highlighting relational, linguistic, and structural patterns that are not evident through manual review alone [35]. In the context of archival records from the trans-Atlantic slave trade, ML-driven pattern recognition allows for the detection of kinship clusters through repeated co-occurrence, approximate name matching, geospatial tagging, and temporal sequencing across fragmented ledgers and plantation inventories [37]. These capabilities are particularly valuable when records were deliberately anonymized, de-individualized, or altered to facilitate commodification. Instead of relying solely on discrete document interpretation, ML enables recognition of continuity across discontinuous sources such as tracking individuals who appear under variant spellings, ages, or labor classifications across different plantation books and tax filings [38].

Another dimension newly visible is the structural imprint of coercion itself. Movement patterns, for instance, can be algorithmically visualized to reveal forced relocations tied to estate consolidation, environmental shocks, or owner indebtedness. Whereas past historiography may have focused on stable plantation units, ML exposes the plantation not as a fixed institution but as a dynamic network of labor redistribution and spatial displacement [36].

In addition, semantic embeddings trained on multilingual corpora help identify possible ethnolinguistic origins of names, prayer songs, or transcription artifacts that clerks themselves may not have recognized as culturally meaningful [39]. This enables the emergence of cultural continuity where the archival surface attempted erasure. However, these forms of visibility are probabilistic, not confirmatory. They illuminate *possibility spaces* clusters of plausibility that guide careful interpretation rather than providing singular factual recovery. Thus, ML transforms historical inquiry not by asserting new certainties, but by expanding what historians know *to look for* [40].

7.2. Constraints and Ethical Limits of Computational Recovery

The same techniques that reveal new relational structures also risk overreach. Algorithms trained on biased data can amplify colonial epistemologies embedded in the archive itself, reproducing the structural hierarchies they aim to critique [36]. Optical character recognition errors, inconsistent orthographies, and clerical distortions introduce uncertainty that must not be mistaken for hidden truth [38]. Moreover, entity resolution across fragmented records can imply kinship or identity continuity where none can be verified.

The ethical limit lies in *interpretive humility*: computational inference must remain provisional, annotated, and reversible [35]. Historians must foreground gaps rather than smoothing them away. Human interpretive judgment remains central, as meaning emerges not from computational correlation but from contextualizing evidence in lived historical worlds [39]. Reparative work cannot rely exclusively on data particularly where data was designed to produce dehumanization. Therefore, computational recovery must resist the impulse to “complete the archive” and instead reveal and honor the structure of its wounds [40].

7.3 Implications for Reparative Historical Memory Work

When used with care, ML supports reparative historiography by restoring *relational presence* rather than presumed certainty [37]. It enables descendant communities, historians, and archivists to collaboratively trace kinship continuities, cultural survivals, and forms of endurance that colonial documentation tried to suppress [36]. The goal is not to overwrite silence, but to acknowledge it, contextualize it, and allow it to speak as evidence of violence and resilience simultaneously [35]. Computational recovery thus becomes a tool for remembrance situating individuals and communities back into history as subjects of meaning rather than commodities of record [40].

8. Conclusion

The central claim of this work is that machine learning does not and cannot replace the interpretive labor of historical analysis. Instead, it expands the field of what can be perceived. ML reveals traces patterns, relationships, recurrences, disruptions but those traces do not speak on their own. They are signs, not meanings. The work of turning them into historical understanding remains with the historian, the community, and the interpretive frameworks that shape how evidence is read and valued.

The archive of the trans-Atlantic slave trade was built to document possession, transfer, and labor extraction not identity, relation, or personhood. Its silences are structural, not accidental. Machine learning enables us to see into those silences differently: not by filling them, but by mapping their patterns, textures, and boundaries. It highlights where names cluster and where they disappear, where lives intersect and where they are forcibly split apart. But even when ML identifies likely kinship connections, shared cultural practices, or spatial continuities, these outputs are forms of *suggestion* rather than verification.

The historian’s role is therefore not diminished it is elevated. The historian must evaluate which computational signals align with cultural practices, with the embodied logic of survival, with the lived rhythms of community. Interpretation requires grounding in social history, oral tradition, and ethical care.

In this model, computational work and historical reasoning are not competing epistemologies they are complementary modes of attention. ML widens the historian's field of vision, making it possible to ask questions we could not have asked before. Historical interpretation then anchors those possibilities in meaning, humanity, and responsibility.

Thus, the task going forward is not to automate recovery, but to deepen collaboration. The aim is not to "solve" the archive but to restore presence, relation, and voice where erasure sought to prevail.

References

[1] Herodotus. *The Histories*. Waterfield R, translator. Oxford: Oxford University Press; 2008.

[2] Ranke L. *The Theory and Practice of History*. Iggers GG, von Moltke K, translators. Indianapolis: Bobbs-Merrill; 1973.

[3] Braudel F. *On History*. Matthews S, translator. Chicago: University of Chicago Press; 1980.

[4] Burke P. *The French Historical Revolution: The Annales School, 1929–1989*. Stanford: Stanford University Press; 1990.

[5] Trouillot MR. *Silencing the Past: Power and the Production of History*. Boston: Beacon Press; 1995.

[6] Stoler AL. *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense*. Princeton: Princeton University Press; 2009.

[7] Eltis D, Richardson D. *Atlas of the Transatlantic Slave Trade*. New Haven: Yale University Press; 2010.

[8] Rediker M. *The Slave Ship: A Human History*. New York: Penguin; 2007.

[9] Berlin I. *Many Thousands Gone: The First Two Centuries of Slavery in North America*. Cambridge (MA): Harvard University Press; 1998.

[10] Johnson W. *Soul by Soul: Life Inside the Antebellum Slave Market*. Cambridge (MA): Harvard University Press; 1999.

[11] Lovejoy PE. *Transformations in Slavery: A History of Slavery in Africa*. Cambridge: Cambridge University Press; 2012.

[12] Fuentes M. *Dispossessed Lives: Enslaved Women, Violence, and the Archive*. Philadelphia (PA): University of Pennsylvania Press; 2016. doi:10.9783/9780812293004

[13] Spillers HT. Mama's baby, papa's maybe: An American grammar book. *Diacritics*. 1987;17(2):65–81.

[14] Sharpe C. *In the Wake: On Blackness and Being*. Durham: Duke University Press; 2016.

[15] McKittrick K. *Demonic Grounds: Black Women and the Cartographies of Struggle*. Minneapolis: University of Minnesota Press; 2006.

[16] Hartman S. *Wayward Lives, Beautiful Experiments: Intimate Histories of Social Upheaval*. New York: W.W. Norton; 2019.

[17] Hartman S. *Lose Your Mother: A Journey Along the Atlantic Slave Route*. New York: Farrar, Straus and Giroux; 2007.

[18] Owens T. *The Theory and Craft of Digital Preservation*. Baltimore: Johns Hopkins University Press; 2018.

[19] Nicholson B, Dawson G. Handwritten text recognition in historical documents: Machine learning approaches and limitations. *Journal of Documentation*. 2019;75(3):511–529.

[20] Atanda ED. 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.;2(12):151-64.

[21] Afolabi Oluwafemi Samson, Femi Adeyemi, Toyyib Oladipo. Effect of transverse reinforcement on the shear behavior of reinforced concrete deep beams. World Journal of Advanced Research and Reviews. 2022;16(2):1294-1303. doi: 10.30574/wjarr.2022.16.2.1267. Available from: https://doi.org/10.30574/wjarr.2022.16.2.1267

[22] Terras M. Digital humanities and digitised cultural heritage. InThe Bloomsbury Handbook to the Digital Humanities 2022 Dec 1 (pp. 255-266). Bloomsbury.

[23] 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.

[24] Eder M, Rybicki J, Kestemont M. Stylometry with R: a package for computational text analysis. R Journal. 2016;8(1):107–121.

[25] 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.

[26] Christen P. Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection. Berlin: Springer; 2012.

[27] 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.

[28] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. Proceedings of ICLR Workshop; 2013. Available from: https://arxiv.org/abs/1301.3781

[29] Kutuzov A, Øvrelid L, Szymanski T, Velldal E. Diachronic word embeddings and semantic shifts: a survey. arXiv preprint arXiv:1806.03537. 2018 Jun 9.

[30] Adeyanju BE, Bello M. Storage stability and sensory qualities of *Kango* prepared from maize supplemented with kidney bean flour and alligator pepper. IOSR Journal of Humanities and Social Science (IOSR-JHSS). 2022;27(1, Series 3):48-55. doi:10.9790/0837-2701034855

[31] Afolabi OS. Load-Bearing Capacity Analysis and Optimization of Beams, Slabs, and Columns. Communication In Physical Sciences. 2020 Dec 30;6(2):941-52.

[32] Eltis D, Richardson D, Roper L, Behrendt S, Klein H. The Trans-Atlantic Slave Trade Database – Voyages. Emory University; 2009. Available from: https://www.slavevoyages.org

[33] Williams D. Digital approaches to the history of the atlantic slave trade. InOxford Research Encyclopedia of African History 2018 Nov 20.

[34] Fett SM. Recaptured Africans: Surviving Slave Ships, Detention, and Dislocation in the Final Years of the Slave Trade. UNC Press Books; 2016 Nov 23.

[35] Dwyer K. Shackles, Collars, and Chains: Exposing the Treatment of Enslaved Black Women during the Middle Passage and as Part of the Archaeological Record (1700–1886). East Carolina University; 2021.

[36] Buchanan RA. Theory and Narrative in the History of Technology. Technology and Culture. 1991;32(2):365-76.

[37] Wyatt S. Technological determinism is dead; long live technological determinism. The handbook of science and technology studies. 2008 Jan 1;3:165-80.

[38] Staudenmaier JM. Technology's storytellers: Reweaving the human fabric. MIT press; 1989 Sep 6.

[39] Smith MR, Marx L, editors. Does technology drive history?: The dilemma of technological determinism. Mit Press; 1994 Jun 2..

[40] Misa TJ. How machines make history, and how historians (and others) help them to do so. Science, Technology, & Human Values. 1988 Jul;13(3-4):308-31.