A Trustworthy and Explainable AI-Assisted Decision Support Framework for WhatsApp Digital Evidence Assessment

Alina Maria Bracho-Rodriguez[1,*], Ciaddy Rodríguez-Borges[1], Jesús Alberto Pérez-Rodriguez[1], Ana Bélen Pérez-Baltar[2], Amanda Isabel Bracho-Rodriguez[3] and Mónica Daniela Barreiro-Linzán[1]

[1]*Universidad Técnica de Manabí, Portoviejo, Ecuador*
[2]*Universidad Politécnica Territorial del Estado Aragua "Federico Brito Figueroa", La Victoria, Venezuela*
[3]*Unidad Educativa Particular Raffaello Santi, Portoviejo, Ecuador*

Abstract

The use of WhatsApp as a source of digital evidence is growing but the forensic analysis of it is difficult since investigators need to maintain integrity, check authenticity, document the chain of custody, and be able to explain their conclusions in a legally defensible manner. Current AI-enabled methods typically focus on problems such as artefact recovery, content analysis, or semantic retrieval rather than the full evidentiary evaluation process. This paper proposes a trustable and explainable AI-assisted decision support framework that combines standardised evidence handling, integrity and metadata checks, source-grounded AI reasoning, an initial Evidence Reliability Index, explainability, privacy safeguards, and mandatory human expert validation. The framework is conceptual and designed to help, not replace, forensic practitioners. In the Ecuadorian judicial framework, an illustrative synthetic scenario explains how its components can interact, without aiming at empirical or legal validation. The proposal provides a structured platform for future implementation, pilot testing and expert evaluation.

Keywords

Digital forensics, WhatsApp, Digital evidence assessment, Artificial intelligence, Explainable artificial intelligence, Trusthy AI, Dectworision support systems, Trustworthy AI, Decision Support Systems

1. Introduction

Digital evidence has become a fundamental source of information in criminal, civil, corporate, and administrative investigations. The widespread use of smartphones and instant messaging applications has transformed digital communications into valuable forensic artifacts that frequently support judicial decision-making. WhatsApp is one of the most thoroughly researched applications among these applications since it preserves chats, multimedia files, documents, voice notes, geolocation records, and metadata that may represent legally significant evidence [1, 2, 3].

There are dedicated forensic tools, yet it is still tough to examine WhatsApp digital data. End-to-end encryption, heterogeneous multimedia formats, erased messages, cloud and multi-device synchronisation and the continual innovation of mobile operating systems hinder acquisition and interpretation. The evidentiary value of an artifact, therefore, depends not only on extraction but also on demonstrating authenticity, integrity, traceability, reproducibility, and procedural compliance before a forensic conclusion can be defended [2, 4, 5].

IWSAI 2026: 1st International Workshop on Systems, Applications, Artificial Intelligence and Innovation, September 16–18, 2026, Piura, Peru
*Corresponding author.
abracho2193@utm.edu.ec (A. M. Bracho-Rodriguez); ciaddy.rodriguez@utm.edu.ec (C. Rodríguez-Borges); jesus.perez@utm.edu.ec (J. A. Pérez-Rodriguez); anabelenperez15@gmail.com (A. B. Pérez-Baltar); brachoamanda21@gmail.com (A. I. Bracho-Rodriguez); monica.barreiro@utm.edu.ec (M. D. Barreiro-Linzán)
0000-0002-5176-9579 (A. M. Bracho-Rodriguez); 0000-0003-1097-4194 (C. Rodríguez-Borges); 0000-0002-1578-2565 (J. A. Pérez-Rodriguez); 0000-0003-3574-1637 (A. B. Pérez-Baltar); 0009-0009-8627-1891 (A. I. Bracho-Rodriguez); 0000-0002-8904-9921 (M. D. Barreiro-Linzán)

Recent advances in Artificial Intelligence (AI) have improved digital forensic investigations. Machine learning, deep learning, natural language processing, Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG) can support evidence retrieval, multimedia analysis, semantic search, and forensic knowledge extraction [6, 7, 8, 9]. At the same time, Explainable Artificial Intelligence (XAI) and Trustworthy AI have become essential in legal environments, where analytical processes must remain transparent, auditable, and subject to human oversight [9, 10, 11, 12, 13, 14].

However, the literature remains fragmented. Most studies address isolated tasks such as artifact acquisition, multimedia authentication, explainability, chain-of-custody management, or intelligent retrieval without bringing these capabilities together in a comprehensive evidence-assessment methodology [8, 10, 15]. In practice, before issuing an expert conclusion, forensic investigators must consider technical characteristics together with contextual consistency, evidentiary relevance, procedural compliance, privacy, and legal reliability.

In this paper digital evidence assessment is seen as a systematic assessment of the authenticity, integrity, contextual consistency, evidential relevance, procedural compliance and legal reliability of digital artefacts in support of forensic decision making. This definition is based on evidence analysis, integrating standardised procedures, legal requirements, expert reasoning and AI-assisted analytical capabilities into a single workflow.

To address the identified gap, this paper proposes a Trustworthy and Explainable AI-Assisted Decision Support Framework for the assessment of WhatsApp digital evidence. Standardised evidence handling and chain-of-custody verification Integrity validation, metadata and contextual analysis AI-assisted reasoning and evidence reliability evaluation Explainability and privacy safeguards Human-centered decision assistance. The use of Ecuador as a context is just illustrative as earlier research found practical and normative shortcomings in the management of WhatsApp evidence in this jurisdiction [16].

This work makes the following main contributions:

- A unified conceptual framework is proposed for complete assessment of WhatsApp digital evidence.
- An initial Evidence Reliability Index is proposed to integrate the technical and procedural dimensions, with no purpose of reaching a decision on legal admissibility.
- Source-grounded AI, explainability, privacy safeguards and mandatory expert validation are part of the forensic workflow.
- A clear development method, an illustrative application scenario situated in Ecuador, and a defined path for future validation are provided.

Figure 1 summarizes the evolution from traditional digital forensics toward explainable and trustworthy decision support, and locates the research gap addressed by this work. The remainder of the paper is organized as follows. Section 2 reviews related work and explains the framework development method. Section 3 presents the architecture. Section 4 provides an illustrative Ecuadorian application scenario. Sections 5 and 6 discuss the contribution and present the conclusions, respectively.

2. Related Work

Artificial Intelligence has transformed digital forensic investigations by expanding traditional evidence extraction towards automated analysis, explainability and decision support. Recent work reports progress in mobile forensics, multimedia authentication, conversation analysis and evidence management. But most contributions remain specialised and are not suited to the combined technical, procedural and legal assessment that is necessary in judicial investigations.

2.1. Artificial Intelligence in Digital Forensics

AI improves evidence acquisition, multimedia analysis, document classification, anomaly detection, and semantic retrieval. Machine learning, deep learning, and transformer-based architectures can reduce

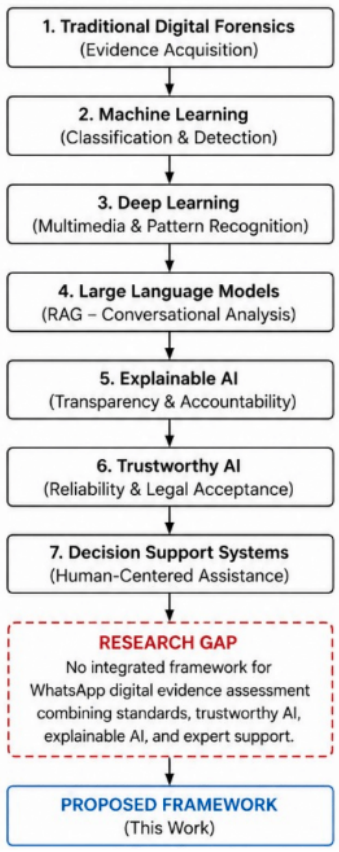

Figure 1: Evolution of AI in digital forensics and the research gap addressed by the proposed framework.

investigation time and improve analytical consistency when they are applied to well-defined forensic tasks [6, 7, 8, 9].

More recently, LLMs and RAG-based interfaces have allowed investigators to explore heterogeneous forensic repositories through natural-language queries. Their usefulness depends on grounding each response in verifiable artifacts and preserving links to the retrieved sources [17].

Despite this progress, current applications remain concentrated on artifact recovery, classification, and content interpretation. Few approaches extend AI support to the complete assessment that precedes an expert conclusion and possible judicial use [9, 10].

2.2. Explainable and Trustworthy Artificial Intelligence

As AI becomes more involved in forensic work, transparency and accountability become essential. XAI provides mechanisms that help investigators understand and justify analytical recommendations [10, 11, 12]. Trustworthy AI adds reliability, robustness, accountability, fairness, privacy, and human oversight, which are particularly important when conclusions must be reproducible and legally defensible [12, 13, 14].

The literature consistently indicates that AI should assist rather than replace forensic practitioners. Human supervision remains necessary to interpret results, identify unsupported inferences, verify procedural compliance, and assume responsibility for the final expert conclusion [13, 14].

2.3. Digital Evidence Reliability and Forensic Standards

The reliability of digital evidence depends on both analytical performance and the integrity of the forensic process. ISO/IEC 27037 guides the identification, collection, acquisition, and preservation of digital evidence, while ISO/IEC 27041 addresses the suitability and adequacy of investigative methods [18, 19].

Related research also highlights chain-of-custody continuity, integrity verification, and standardized documentation as conditions that strengthen evidentiary confidence [20, 21].

These contributions show that evidentiary value cannot be derived by extraction or model accuracy alone. This is due to the combined application of sound technical procedures, documented procedures, contextual interpretation and expert review.

2.4. Gap in the Literature

We have made significant progress in AI-assisted digital forensics, XAI, trustworthy AI, forensic standards and decision support. However, these lines of research are mostly developed separately. Existing studies commonly address one or two stages of the forensic lifecycle without integrating acquisition, integrity verification, contextual analysis, evidence reliability, explanation, and expert validation within a single methodology [9, 13, 15, 22, 23, 20, 21].

Research on WhatsApp evidence has mainly focused on artifact recovery, encrypted databases, metadata extraction, and legal use [1, 2, 3, 4, 5, 8]. These contributions are valuable, but they provide limited support for assessing evidentiary reliability across the complete lifecycle. The remaining gap is therefore not the absence of individual techniques, but the absence of a coordinated and human-centered assessment process.

The proposed framework responds to this integration gap. It does not claim that previous approaches are technically inferior; instead, it combines their complementary strengths into a structured architecture for assessing WhatsApp evidence.

2.5. Framework Development Method

The framework was constructed through a structured narrative synthesis rather than a new systematic review. The starting point was the authors' prior PRISMA-based evaluation of WhatsApp digital evidence for the period 2020-2025 that searched Scopus, IEEE Explore and Google Scholar and found recurrent technological and legal constraints [16]. This was followed by a focused update of recent work on AI-assisted digital forensics, XAI, trustworthy AI, decision support, chain of custody and forensic standards.

We focused on peer-reviewed studies and official standards, primarily published between 2021 and 2026, but included older foundational work when it directly assisted decision support or explainability. The search terms included combinations of the terms "digital forensics," "artificial intelligence," "explainable AI," "trustworthy AI," "decision support," "WhatsApp," "chain of custody," "ISO/IEC 27037," and "ISO/IEC 27041." Duplicates, sources with no apparent forensic function and studies not related to the assessment of evidence were excluded.

Next, the selected requirements were mapped to the evidence lifecycle functional modules: acquisition, preservation, validation, interpretation, reliability assessment, explanation, expert review, and reporting. The qualitative synthesis in Table 1 summarizes the main research streams used to derive the framework. It is intended to show complementarity and remaining integration needs, not to rank individual studies.

The synthesis in Table 1 shows that the literature provides strong components for individual stages of digital forensics, but limited integration across the whole assessment process. Hence the novelty claimed in this paper is architectural and methodological: the proposed framework unifies these complementary components while explicitly maintaining human responsibility and acknowledging that the empirical validation remains to be done.

Table 1
Qualitative synthesis of the research streams supporting the proposed framework.

Research stream	Established contribution	Remaining limitation
AI-based digital forensics [6]–[9]	Automates extraction, classification, semantic retrieval, and multimedia analysis.	Usually addresses individual technical tasks rather than the full evidentiary assessment.
XAI and Trustworthy AI [9]–[14]	Improves interpretability, transparency, reliability, and human oversight.	Often evaluated at model level and not connected to evidence handling or chain of custody.
Forensic standards [18], [19]	Defines evidence handling and method-assurance requirements.	Does not provide AI reasoning or a WhatsApp-specific assessment workflow.
Decision-support approaches [15]–[23]	Structures investigative reasoning, prioritization, and reporting.	Not specifically designed for WhatsApp evidence or integrated with XAI and standards.
Chain-of-custody and integrity methods [20], [21]	Strengthens traceability and helps detect unauthorized changes.	Frequently focuses on a particular artifact type or technical mechanism.
WhatsApp digital forensics [1]-[5], [8]	Supports recovery and examination of messages, databases, media, and metadata.	Provides limited support for integrated reliability assessment and expert-centered decision making.

3. Trustworthy and Explainable AI-Assisted Decision Support Framework

3.1. Framework Overview

The proposed framework supports the comprehensive assessment of WhatsApp digital evidence by integrating standardized procedures, AI, XAI, trustworthy AI principles, privacy safeguards, and human-centered decision support. Unlike approaches that automate only one task, the architecture follows the evidence from acquisition to the preparation of a structured forensic report.

The framework is founded on three principles. First, digital evidence must be processed according to documented and reproducible procedures. Secondly, AI is an analytical assistant, not a gatekeeper of legal admissibility. Third, each AI generated observation should be traceable to its source artefacts, should include information about uncertainty and should be reviewed by a qualified forensic practitioner.

It starts with gathering legal evidence and ends with a report classifying verified facts, analytical observations, limitations and expert opinions. Standards, integrity checks, metadata analysis, AI reasoning, reliability assessment, and human supervision interact throughout the process. The overall architecture is shown in Figure 2.

Each module performs a defined forensic function and exchanges structured information with the next stage. This modular design may support future integration with existing forensic platforms and allows individual analytical components to be updated without changing the complete workflow.

3.2. Evidence Acquisition and Standardized Evidence Handling

The framework begins with the lawful acquisition of WhatsApp conversations, databases, multimedia files, documents, backups, notification records, and related metadata. The acquisition record should identify the legal authority, device state, examiner, date and time, tool and version, extraction method, and generated hash values. These activities follow the handling principles of ISO/IEC 27037, while the selected tools and methods should be shown to be fit for purpose under ISO/IEC 27041 [18, 19].

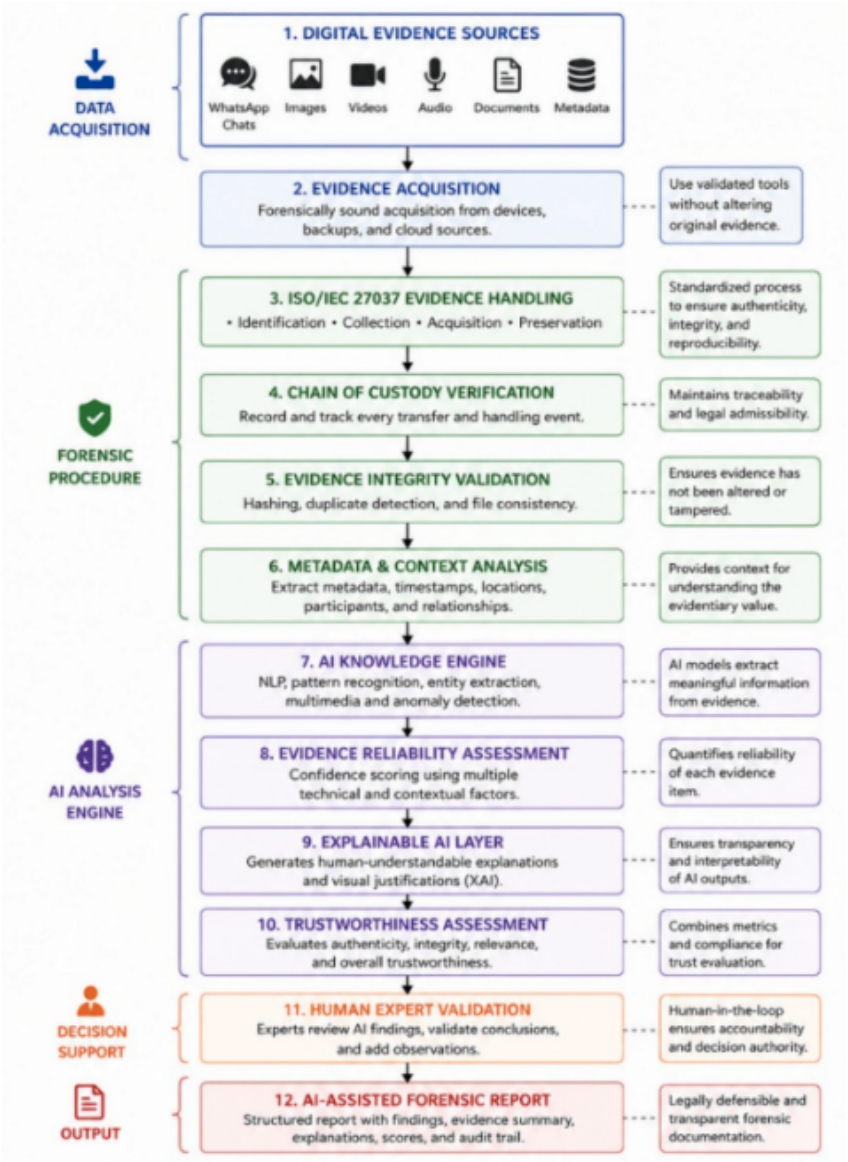

Figure 2: A Trustworthy and Explainable AI-Assisted Decision Support Framework for WhatsApp Digital Evidence Assessment.

The framework is not intended to bypass end-to-end encryption. It checks for content that is lawfully on the device, associated backups or other permitted sources. If encrypted content is not available this is logged as a limitation, and not reconstructed by A.I. Deleted artefacts can be searched for in database remnants, notification logs, caches or backups, but any recovered fragments are explicitly identified and require corroboration before they can influence an expert opinion.

Integrity is verified through cryptographic hash, acquisition logs, file and database consistency checks, and chain-of-custody records, before AI-assisted analysis. In the case of data that is multi-device or cloud-synchronized, the examiner will check time stamps, device identifiers, message identifiers, file hashes and synchronisation records. Conflicts remain as uncertainty and are not automatically resolved.

3.3. Context Assessment and Metadata

After verifying integrity, the framework parses and structures timestamps, device identifiers, communication data, sender-recipient relationships, geolocation data, multimedia properties and application-specific metadata. Images, music, video and documents still remain linked to the source from whence

they were extracted and the original message.

The goal is to rebuild context, instead than treating metadata as a collection of separate qualities. Comparing across sources helps uncover communication sequences, duplicated or missing records, discrepancies in synchronisation, and links between messages and multimedia artefacts. Contextual contradictions are not concealed, but noted as variables that could diminish evidence trustworthiness.

3.4. AI Knowledge Engine

The analytical core of the framework is the AI Knowledge Engine. It can organise evidence by the combination of natural language processing, multimedia analysis, semantic retrieval and pattern recognition, and it can help interpretation. Its only function is to generate reviewable observations and investigatory hypotheses.

The engine offers four complementary functions:

- Semantic text interpretation;
- Organization and assisted analysis of multimedia contents;
- Recognition of patterns in timeline and context;
- Developing evidence packages for expert review.

Queries are limited to the validated case repository regardless of whether it is LLMs or RAG. All responses should contain the supporting artefact or retrieved passage, note the model and prompt version, and distinguish between source-based observations and created interpretation [17]. Retrieval and possible bias of contradicting sources, poor confidence in retrieval, and statements made without proof are marked for human review. This management minimises the danger of hallucination, but not its elimination.

3.5. Evidence Reliability Assessment

A distinctive component of the framework is the Evidence Reliability Assessment module. Unlike a conventional model-confidence value, it evaluates the condition of the evidence and the quality of the forensic process.

For a preliminary operationalisation, six dimensions are scored from 0 to 1 based on documented checks:

- Authenticity (A; weight 0.20): source, account, device and artefact identifier consistency;
- Integrity (I; weight 0.25): effective verification of the hash and no unexpected changes;
- Contextual consistency (C; weight 0.15): consistency of the chronology, participants and related artefacts;
- Metadata completeness (M; weight 0.10): existence of the required technical attributes;
- Procedural compliance (P; weight 0.15): documented acquisition, tool use and technique assurance;
- Chain-of-custody continuity (CoC; weight 0.15): full and traceable history of evidence handling.

First Evidence Reliability Index is determined as follows: ERI = 0.20A+0.25I+0.15C+0.10M+0.15P+0.15CoC Reliability is high when the values are between 0.80 and 1.00, moderate when the values are between 0.60 and 0.79 and low when the values are less than 0.60.

A dimension may also be marked “not assessable” when information is unavailable. The score is an auditable decision-support indicator, not a legal ruling. The expert may modify the interpretation only by recording a written justification.

3.6. Explainable AI and Human Expert Validation

Every AI-assisted observation is accompanied by an explanation that identifies the supporting artifacts, relevant metadata, analytical relationship, uncertainty, and any conflicting information. The framework therefore emphasizes local and case-specific explanations rather than a general statement that the model is accurate.

Explainability may later be evaluated through source-traceability rate, explanation completeness, expert agreement, and perceived usefulness. Recommendations that are not supported by verified evidence are excluded from the reliability assessment and cannot be reported as a confirmed finding. Trustworthiness, in this context, is a crosscutting evaluation that combines the quality of evidence, explanatory power, privacy and security precautions, and monitoring by humans before a finding is included in the final report.

Human expert validation is needed. The examiner reviews evidence, AI-assisted observations, explanations, procedural records, and reliability indications before accepting, changing or rejecting a suggestion. Final responsibility remains with the forensic professional, preserving judicial accountability and preventing the framework from functioning as an autonomous decision-maker.

3.7. Privacy, Data Security and Ethical Safeguards

WhatsApp evidence may include private conversations, geolocation data, identifiers, and sensitive multimedia content. Privacy and ethical safeguards thus apply across all modules. The framework calls for purpose limitation, data minimisation, role-based access, encryption at rest and in transit, activity logging, controlled retention and redaction of information that is not relevant to the investigation.

In the Ecuadorian context, these controls should be aligned with the Organic Law on Personal Data Protection, particularly its principles of security, proportionality, confidentiality, and lawful processing [24]. AI models should not be trained on case data unless a valid legal basis, adequate safeguards, and clear governance conditions exist. Table 2 summarizes the functional components and cross-cutting safeguards of the framework.

3.8. Forensic Report Generation with AI

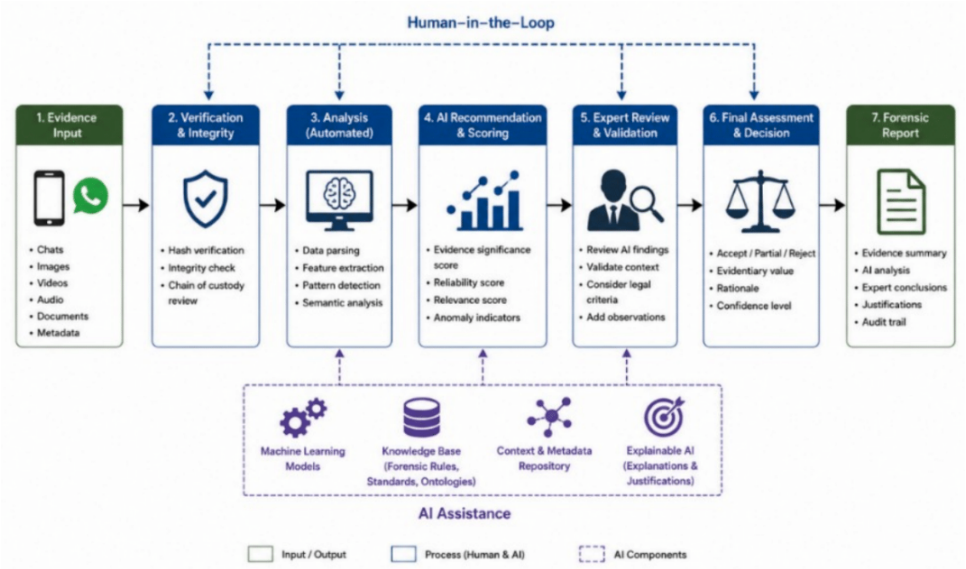

Figure 3: AI-assisted digital evidence assessment workflow.

The last module generates a forensic report in a structured format, integrating the acquisition records, integrity results, AI-assisted observations, evidence reliability indications, privacy restrictions, and

Table 2
Functional components of the proposed framework.

Module or safeguard	Primary function
Evidence Acquisition	Lawful collection of WhatsApp artifacts and associated records.
Standardized Evidence Handling	ISO/IEC 27037 handling and ISO/IEC 27041 method assurance.
Chain of Custody	Documented traceability of every transfer and forensic action.
Integrity Validation	Hash verification and consistency checks before AI analysis.
Metadata and Context Analysis	Reconstruction of timelines, relationships, and cross-source context.
AI Knowledge Engine	Source-grounded organization and assisted interpretation of evidence.
Evidence Reliability Assessment	Weighted evaluation of six technical and procedural dimensions.
Explainable AI	Artifact-level explanations, uncertainty, and conflict reporting.
Trustworthiness Assessment	Cross-cutting review of reliability, explainability, privacy and security safeguards, and human oversight before reporting.
Human Expert Validation	Acceptance, modification, or rejection of AI recommendations.
Privacy and Ethical Safeguards	Data minimization, access control, security, retention, and accountability.
Structured Forensic Report	Separation of verified facts, observations, limitations, and expert conclusions.

the expert opinions. It does not resolve the case. The paper instead clearly distinguishes confirmed findings, recovered or incomplete artefacts, analytical hypotheses, disputed data and constraints. The operational sequence is shown in Figure 3.

The workflow shows that AI analysis occurs only after acquisition, preservation, chain-of-custody documentation, and integrity verification. Reliability assessment, explanation, and human validation then operate as safeguards before the final report is issued.

4. Illustrative Application in the Ecuadorian Judicial Context

To illustrate how the modules interact without claiming empirical validation, consider a synthetic scenario based on recurring issues identified in Ecuadorian research [16]. A WhatsApp chat, photographs, voice notes and gaps that may have occurred from deletion or multi-device syncing, on a lawfully seized Android phone. Before doing any AI analysis, the examiner documents the legal authority, device state, extraction procedure, tool version, timestamps, hash values, and any custody transfer.

It doesn't circumvent end-to-end encryption. The framework will only use data that is legally available on the device, associated backups or other allowed sources. Recovered fragments stored in databases, caches or notification logs are tagged as recovered and are assigned lower completeness or contextual scores unless corroborated. Timestamps, message IDs, device records and file hashes are compared to identify synchronisation conflicts or unexplained changes.

The AI Knowledge Engine structures participants, events, media, and relationships in time, while RAG queries are still limited to the verified case repository. Each observation points to its source artifact and is classified as a verified fact, analytical observation, or hypothesis. The ERI combines technical and procedural dimensions, and the forensic expert may accept, modify, or reject the recommendation

before the structured report is produced.

In Ecuador, the processing of private communications must also respect purpose limitation, proportionality, security, confidentiality, and the rights established in the Organic Law on Personal Data Protection [24]. This scenario demonstrates the intended use of the framework, but it is not a real case study, a software implementation, or proof of judicial admissibility. No formal validation with Ecuadorian forensic practitioners, prosecutors, or procedural-law specialists was conducted during this conceptual design phase; that consultation is part of the planned validation pathway.

5. Discussion

Because the proposed framework is conceptual in nature, this section discusses its expected contribution, limitations and validation needs instead of reporting empirical performance results. The main contribution is the integration of processes usually handled separately: evidence handling, integrity, AI-assisted analysis, reliability assessment, explanation, privacy, expert review and reporting.

ISO/IEC 27037 and ISO/IEC 27041 define the procedures for handling evidence and providing assurance of the methods [18, 19]. Decision-support studies provide structured investigative reasoning [15, 22]. Chain-of-custody studies provide traceability and integrity controls [20, 21]. The framework does not replace those approaches but coordinates them so that the AI-generated observations are only produced once the evidence and method are documented and checked.

The preliminary ERI gives more concreteness to the reliability module by making its dimensions, weights, and thresholds explicit. Nevertheless, the suggested weights are assumptions for the design and are not to be interpreted as universally valid. Their stability needs to be tested on various cases, tools, platforms and expert teams. Legal admissibility is also a judicial finding that is not quantifiable to a numerical score.

LLMs bring with them new hazards. Even source grounded systems may produce partial, biased or inaccurate interpretations. Thus, the framework needs artifact-level citations, logging of models and prompts, reporting of uncertainty, detection of conflicts and human approval. Privacy hazards are further mitigated by data minimisation, limited access, safe storage, and the prohibition of secondary model training without a suitable legal basis and governance mechanism.

A first validation pilot should be based on a synthetic and fully documented WhatsApp dataset, including messages, multimedia, deleted fragments, synchronisation conflicts and known chain-of-custody events. Forensic practitioners could compare the framework against their normal workflow when used in a laboratory setting. Relevant measures include task completion time, artefact traceability, inter-expert agreement, ERI stability, source-grounding rate, hallucination frequency, explanation usefulness, and the percent of AI recommendations accepted or corrected by experts. A later phase should make use of real case files, anonymised, and with institutional and legal authorisation.

6. Conclusion

This paper proposed a Trustworthy and Explainable AI-Assisted Decision Support Framework for WhatsApp digital evidence assessment. The architecture was derived from a structured synthesis of prior WhatsApp research, recent AIassisted forensic studies and the requirements of ISO/IEC 27037 and ISO/IEC 27041. It connects lawful acquisition, integrity and chain-of-custody verification, contextual analysis, source-grounded AI reasoning, reliability assessment, privacy safeguards, expert validation, and structured reporting.

The paradigm transforms the discourse away from individual AI applications to a thorough and auditable process of evidence assessment The preliminary ERI offers an operational foundation for the evaluation of authenticity, integrity, context, metadata, procedural compliance and custody continuity. The following Ecuadorian situation demonstrates how these variables may be assessed in light of the jurisdictions.

AI is still a support tool and the forensic practitioner is responsible for interpretation and reporting. The primary limitation is the conceptual nature of the proposal. Future work should involve building a proof of concept, running a lab evaluation with forensic experts, calibrating the ERI, measuring explanation quality and hallucination risk, and later considering anonymised real cases and other messaging platforms.

Acknowledgments

The authors thank the Integrated, Sustainable and Intelligent Engineering Research Group (GI ISI) of the Universidad Técnica de Manabí and the Ibero-American Network for the Large-Scale Integration of Renewable Energy into Electric Power Systems (RIBIERSE-CYTED) for their academic and technical support.

Declaration on Generative AI

During the preparation of this paper, ChatGPT (OpenAI) was used for language refinement, editorial consistency, and organization of revisions requested during peer review. All scientific claims, references, interpretations, and final wording were reviewed and approved by the authors, who take full responsibility for the manuscript.

References

[1] S. A. E. Sarhan, H. A. Youness, A. M. Bahaa-Eldin, A framework for digital forensics of encrypted real-time network traffic, instant messaging, and voip application case study, Ain Shams Engineering Journal 14 (2023) 102069.

[2] R. Adijisman, I. Riadi, Mobile forensic on whatsappservices using national institute of standards and technology method, International Journal of Computer Applications 183 (2021) 41–48.

[3] S. N. Aniza, I. Riadi, Mobile forensic for cyber fraud case on whatsapp services using national institute of standard technology method, International Journal of Computer Applications 183 (2021) 44–50.

[4] D. Banegas-Crespo, D. Andrade-Pesantez, Análisis forense en dispositivos móviles android para casos de ciberextorsión, revisión sistemática de literatura. mqrinvestigar, 8 (3), 4076–4097, 2024.

[5] S. A. A. Lema, A. J. A. Lema, M. A. D. Correa, La prueba digital en el proceso penal ecuatoriano, desafíos constitucionales, técnicos y comparados frente a los estándares iberoamericanos de evidencia tecnológica, Ciencia Latina Revista Científica Multidisciplinar 9 (2025) 17880–17900.

[6] J. M. Á. Silva, O. G. Rodríguez, J. A. Durán, Facebook y whatsapp como pruebas digitales en el sistema penal acusatorio en méxico, Advocatus (2023) 57–75.

[7] C.-E. Bogos, R. Mocanu, E. Simion, A security analysis comparison between signal, whatsapp and telegram, Cryptology ePrint Archive (2023).

[8] T. A. Cahyanto, M. A. Rizal, A. E. Wardoyo, T. T. Warisaji, T. Timur, Live forensic to identify the digital evidence on the desktop-based whatsapp, Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) 6 (2022) 213–219.

[9] S. Alam, Z. Altiparmak, Xai-cf—examining the role of explainable artificial intelligence in cyber forensics, Engineering Applications of Artificial Intelligence 167 (2026) 113892.

[10] L. Kelly, S. Sachan, L. Ni, F. Almaghrabi, R. Allmendinger, Y.-W. Chen, Opportunities, challenges and a drug testing case study, Digital Forensic Science (2020) 159.

[11] P. Hermosilla, S. Berríos, H. Allende-Cid, Explainable ai for forensic analysis: a comparative study of shap and lime in intrusion detection models, Applied Sciences 15 (2025) 7329.

[12] J. M. Durán, D. van der Vloed, A. Ruifrok, R. J. Ypma, From understanding to justifying: Computational reliabilism for ai-based forensic evidence evaluation, Forensic Science International: Synergy 9 (2024) 100554.

[13] H. Qian, L. Xia, R. Ge, Y. Fan, Q. Wang, Z. Jing, From black boxes to glass boxes: Explainable ai for trustworthy deepfake forensics, Cryptography 9 (2025) 61.
[14] I. Hefetz, Mapping ai-ethics' dilemmas in forensic case work: To trust ai or not?, Forensic science international 350 (2023) 111807.
[15] G. Horsman, Formalising investigative decision making in digital forensics: Proposing the digital evidence reporting and decision support (derds) framework, Digital Investigation 28 (2019) 146–151.
[16] A. M. Bracho-Rodriguez, R. Toala-Dueñas, Evidencia digital de whatsapp en la justicia ecuatoriana: Revisión sistemática (2020-2025), Revista Científica INGENIAR: Ingeniería, Tecnología e Investigación. ISSN: 2697-3693. 9 (2026) 510–536.
[17] P. Praveen, A. Krishna, An ai-powered conversational system for rapid digital forensic analysis, 2026.
[18] I. O. for Standardization ISO, Information technology-Security techniques-Guidelines for identification, collection, acquisition, and preservation of digital evidence, na, 2012.
[19] International Organization for Standardization ISO, Information technology—security techniques—guidance on assuring suitability and adequacy of incident investigative method, 2015. URL: https://www.iso.org/standard/44405.html.
[20] H. M. Elgohary, S. M. Darwish, S. M. Elkaffas, Improving uncertainty in chain of custody for image forensics investigation applications, IEEE Access 10 (2022) 14669–14679.
[21] M. Ali, A. Ismail, H. Elgohary, S. Darwish, S. Mesbah, A procedure for tracing chain of custody in digital image forensics: A paradigm based on grey hash and blockchain, Symmetry 14 (2022) 334.
[22] A. Nisioti, G. Loukas, A. Laszka, E. Panaousis, Data-driven decision support for optimizing cyber forensic investigations, IEEE Transactions on Information Forensics and Security 16 (2021) 2397–2412.
[23] S. Habibi-Chenaran, B. Samadirad, A. T. Miandoab, P. Rezaei-Hachesu, T. S. Soltani, A decision support system to determine the amount of wergild and compensation based on forensic medicine clinical examinations, Egyptian Journal of Forensic Sciences 14 (2024) 10.
[24] Asamblea Nacional del Ecuador, Ley orgánica de protección de datos personales, 2021. URL: https://www.registroficial.gob.ec/quinto-suplemento-al-registro-oficial-no-459/.