International Journal of Law Analytics (IJLA)
Vol. 4, No. 4 2026: 599-614

Beyond Admissibility: Reconstructing Legal Validity and Algorithmic Reliability Standards for AI-Based Electronic Evidence in Indonesian Criminal Procedure

Aloysius Wisnubroto[1*], Iwan Anggoro Warsita[2]
Universitas Atma Jaya Yogyakarta
Corresponding Author: Aloysius Wisnubroto al.wisnubroto@uajy.ac.id

A R T I C L E I N F O

Keywords: AI-Based Electronic Evidence, Criminal Procedure, Admissibility, Algorithmic Reliability, Due Process

Received : 09, July
Revised : 11, September
Accepted: 13, November

©2026 Wisnubroto, Warsita:
This is an open-access article distributed under the termsof the Creative Commons Atribusi 4.0 Internasional.

A B S T R A C T

This article examines the evidentiary gap between procedural legality and algorithmic reliability in Indonesia following Law Number 20 of 2025 on the Criminal Procedure Code. Using normative legal research, statutory interpretation, conceptual analysis, and comparative examination of State v. Loomis and State v. Miles, it identifies limits in applying authentication and lawful-acquisition requirements to AI-generated conclusions. The study proposes a two-layer assessment framework combining procedural legality, authenticity, integrity, and custody with contextual technical validation, error analysis, transparency, and independent testing. It argues that proportionate disclosure, meaningful defense challenges, and reasoned judicial evaluation are essential to preventing algorithmic scores from acquiring excessive probative weight. The framework offers a prescriptive foundation for judicial guidance, not an existing statutory audit requirement.

DOI: https://doi.org/10.59890/ijla.v4i4.374
E-ISSN: 3032-4041
https://slamultitechpublisher.my.id/index.php/ijla

INTRODUCTION

Digital transformation has changed both the production of criminal evidence and the means by which legal actors interpret it. Investigators increasingly encounter surveillance video, device records, communications metadata, biometric information, and other digital traces. Artificial intelligence (AI) adds a further layer: computational systems no longer merely preserve information but classify, enhance, compare, or infer facts from it. Facial-recognition software may generate candidates from a poor-quality image; a speech model may compare recordings; and a detection model may assign a likelihood that audiovisual content was manipulated. Such outputs may look precise, yet the numerical appearance of certainty does not establish the reliability or legal meaning of an algorithmic conclusion (Gless, 2020; Roth, 2016).

This transformation has particular significance for criminal procedure because evidentiary errors can affect liberty, reputation, and the accuracy of judicial decisions. A lawfully collected video may be authentic while a facial-recognition match derived from it is unreliable. Conversely, a technically sophisticated analysis may rest on data acquired through an unlawful search. Neither circumstance can be understood if the law treats the existence of an electronic file, the legality of its acquisition, and the accuracy of an inferential model as one undifferentiated question. Criminal adjudication requires separate answers concerning procedural legitimacy, technical trustworthiness, relevance, and the probative value ultimately assigned by a court (Gless et al., 2024).

Indonesia's Law Number 20 of 2025 on the Criminal Procedure Code, effective from 2 January 2026, gives this issue contemporary urgency. Article 235(1)(f) expressly includes electronic evidence among legally recognized evidentiary categories. Article 235(3) requires evidence to be authenticated and obtained without violating the law; Article 235(4) authorizes judicial scrutiny of those matters, and Article 235(5) states the consequences when evidence is found unauthentic or unlawfully obtained. Article 242 encompasses electronic information, electronic documents, and electronic systems connected to criminal acts. These provisions establish meaningful safeguards, but they do not themselves prescribe a detailed method for evaluating model validity, error distributions, training-data limitations, or the reproducibility of AI-assisted analyses (Republic of Indonesia, 2025).

The Electronic Information and Transactions Law, as amended, already recognizes electronic information, documents, and printouts as lawful evidence under its applicable conditions. The Personal Data Protection Law also raises questions about biometric data and the handling of information belonging to individuals who may not be defendants. The coexistence of these rules does not eliminate the need to distinguish their functions: the legality of obtaining data, the trustworthy operation of a system, and the weight of an inference are related but nonidentical matters. A harmonized criminal-procedural approach must protect the admissibility boundary without letting formal compliance conceal a scientifically weak result (Ramiyanto, 2017; Wachjoe P., 2016).

Current scholarship has established important foundations for this inquiry. Gless (2020) explains why machine-generated evidence challenges procedures historically designed around human accounts. Roth (2016) identifies the risks of transferring evidentiary authority to computational systems that neither testify nor explain themselves as human witnesses do. Technical scholarship reveals how algorithmic opacity, discriminatory error rates, and contextual performance complicate apparently objective outputs (Barocas & Selbst, 2016; Burrell, 2016). The National Institute of Standards and Technology's AI Risk Management Framework likewise treats validity, reliability, transparency, and risk assessment as context-dependent dimensions of trustworthy technology (Tabassi, 2023). However, these strands are often discussed separately rather than integrated into a workable examination sequence under Indonesia's new procedural code.

Comparative judicial developments provide a useful, though limited, perspective. In State v. Loomis (2016), the Wisconsin Supreme Court addressed the constrained use of a proprietary risk assessment in sentencing, not the trial admissibility of facial-recognition evidence. More directly, State v. Miles (2026) concerned the information a defendant should receive about facial-recognition technology used during a criminal investigation. The New Jersey Supreme Court required disclosure about the tools used and how they were deployed while declining automatically to compel disclosure of proprietary source code without a further showing of particularized need. These decisions are not binding precedents in Indonesia, but they clarify the importance of human judgment, tailored disclosure, and meaningful defense challenges.

The central research gap therefore concerns the relationship between an evidence item's legally permissible entry into criminal proceedings and the trustworthiness of an AI-generated conclusion derived from that item. This article proposes a two-layer assessment framework: procedural legality and authenticity form the first layer, while technical reliability and contextual validity form the second. The framework is complemented by disclosure, adversarial testing, expert scrutiny, and independent judicial evaluation of probative value. Its novelty lies not in proposing AI output as an autonomous category of evidence but in giving courts a disciplined method to prevent the legality of an electronic record from becoming a proxy for scientific certainty.

This study asks how Indonesia's current regulation of electronic evidence leaves gaps in assessing AI-generated information and how a legally grounded, technically informed framework can address them while safeguarding defendants' rights. It examines the relevant procedural and electronic-information legislation, develops distinctions between authenticity and inferential reliability, considers lessons from two foreign judgments, and formulates an operational model for Indonesian criminal adjudication. The objective is a normative proposal rather than an empirical claim that Indonesian courts already apply the recommended technical criteria.

THEORETICAL REVIEW

Electronic Evidence and the Limits of Formal Admissibility

Electronic evidence is not a homogeneous type of information. A file may be a direct digital record of an event, an output from a sensor, an edited representation, or an inference produced by a model trained on data that do not appear in the case file. These origins matter because different claims require different justification. Authenticating a video may establish that a particular file came from a stated camera and has not been impermissibly changed. It does not by itself establish that a person depicted in the video has been correctly identified by facial-recognition software. Gless et al. (2024) emphasize the need to adapt evidence law to information generated or processed by computational systems rather than simply treating machines as reliable substitutes for witnesses.

Admissibility, authenticity, and evidentiary weight must therefore be distinguished analytically. Admissibility concerns whether the legal system permits an item to be considered; authenticity concerns whether the item is what its proponent claims; and weight concerns how strongly it supports a contested factual proposition. Technical reliability adds a further inquiry when an algorithm converts an observation into an estimate, score, ranking, or prediction. The final inference remains the responsibility of the fact-finder. Scientific validity cannot be reduced to a product reputation, and a computer-generated score cannot independently establish a statutory element of an offense (Roth, 2016; Murphy, 2007).

Algorithmic Reliability, Explainability, and Contestability

An AI model's reliability is relational: it depends on its design, data, testing conditions, and particular use. A facial-recognition system can perform well on one benchmark yet encounter materially different error patterns with low-resolution footage, unusual lighting, or a database unlike the original test set. NIST's evaluation of facial-recognition systems illustrates why performance claims must be read against the relevant task and demographic or image conditions (Grother et al., 2019). False matches can implicate innocent people, while missed matches can misdirect or delay investigations; neither risk is adequately described by a single headline accuracy figure.

Opacity and contestability are equally central. Proprietary or complex models may prevent an affected person from understanding how a result was produced, but explanation need not always require public release of every line of source code. Depending on the disputed issue, model identity, version, performance documentation, input quality, processing logs, selection thresholds, and an expert's reasoned account may provide a meaningful basis for challenge. The key normative question is whether the defense can test a material proposition that the prosecution seeks to prove, not whether absolute algorithmic transparency is technologically possible (Burrell, 2016; Adadi & Berrada, 2018; Rudin, 2019).

Due Process and Judicial Epistemic Responsibility

A fair criminal proceeding requires that an accused person can challenge material adverse information. Algorithmic evidence creates a potential asymmetry when investigators or vendors control the data and technical information necessary to assess error. The resulting disadvantage is not solely a technological problem: it concerns equality of arms, access to relevant evidence, and the reasons a court gives for relying on specialist testimony. Citron and Pasquale (2014) identify procedural concerns raised by automated assessments, while Roth (2016) cautions against treating machine outputs as epistemically self-proving.

The proposed conceptual framework therefore links three propositions. First, legality and authentication are threshold inquiries governed by positive law. Second, reliability is a contextual appraisal of the method actually used to produce the asserted result. Third, judicial justification must explain how the particular result relates to the charged conduct and other evidence. The first two inquiries are complementary rather than interchangeable, and the third prevents a technically credible but legally irrelevant result from acquiring undeserved weight. This sequence is a normative analytical design developed here; it should not be mistaken for a technical audit obligation already enacted verbatim in Indonesian legislation.

METHODOLOGY

This article uses normative legal research with statutory, conceptual, comparative, and case approaches. The statutory analysis examines Law Number 20 of 2025 on the Criminal Procedure Code, the Electronic Information and Transactions Law and its amendments, and associated rules relevant to digital systems and personal information. The constitutional perspective draws on the guarantees of fair legal certainty and personal protection in Articles 28D (1) and 28G (1) of the 1945 Constitution. The inquiry evaluates the legal significance of authentication and lawful acquisition while identifying technical questions left for further elaboration in practice.

The conceptual approach draws on legal scholarship concerning machine evidence, forensic validity, algorithmic opacity, and procedural fairness, together with NIST guidance on AI risk management. The comparative and case approaches examine State v. Loomis (2016) and State v. Miles (2026) for their reasoning about computational tools, disclosure, and the limits of judicial reliance. Because the decisions arose in different legal settings and procedural stages, the analysis extracts transferable considerations rather than treating their holdings as Indonesian rules. Materials were examined doctrinally through textual, systematic, and purposive interpretation, followed by synthesis into a prescriptive model.

The study does not collect interview, survey, laboratory, or court-frequency data. Accordingly, its findings are normative propositions about legal design and institutional practice rather than statistical estimates of AI error in Indonesia. Technical publications are used to identify relevant validation questions, not to certify the performance of any specific software product. This methodological boundary is important: the proposed framework can structure judicial inquiry, but the accuracy of a particular model must be demonstrated with case-appropriate evidence rather than assumed from the literature.

RESEARCH RESULTS

Normative Recognition of Electronic Evidence Under the New Criminal Procedure Code

The first result is that Indonesia's legal framework now expressly recognizes electronic evidence while continuing to place legal conditions on its use. Article 235(1)(f) of Law Number 20 of 2025 makes electronic evidence an enumerated evidentiary category, and Article 242 defines its reach by reference to electronic information, electronic documents, or electronic systems connected to an offense. Article 235(3) requires authentication and lawful acquisition, and Article 235(5) excludes evidence found unauthentic or unlawfully acquired from use in court and denies it probative force. This structure answers whether an electronic item may be legally examined, but it does not provide a comprehensive scientific test for an algorithmic interpretation of that item (Republic of Indonesia, 2025).

The distinction is most visible when AI is used only after lawful collection. Suppose a surveillance file is secured through an authorized process, preserved in its original form, and supported by appropriate records of possession. Those measures may provide strong reasons to accept the file's provenance. Yet if the prosecution then invokes a recognition system to label a face, the factual assertion at issue changes. The question is no longer only whether the original video existed and remained intact; it is whether the model correctly associates the observed image with an individual under the actual conditions of capture. Legal compliance at acquisition is necessary but is not a substitute for validation of the latter inference.

The Electronic Information and Transactions Law recognizes electronic information and documents within its own normative framework. Government Regulation Number 71 of 2019 addresses electronic system administration, and Law Number 27 of 2022 provides a broader personal-data protection framework. These instruments supply related obligations and concepts, but their distinct purposes should be preserved. Compliance with duties concerning an electronic system does not automatically establish the truth of a generated identification; nor does technical validity erase questions about an unlawful intrusion into a person's private data. The first doctrinal gap is therefore not the total absence of regulation but the absence of an expressly integrated evaluation standard for AI-mediated evidentiary claims.

Procedural Weaknesses: Acquisition, Provenance, and Transformation Records

A second finding concerns the limits of conventional documentation when records are transformed by AI. A reliable chain of custody identifies who collected, copied, transferred, stored, and examined electronic material. Investigators should preserve the original source where feasible, document acquisition methods, and record integrity checks such as cryptographic hashes. These practices support provenance and help reveal unexplained alteration, but they must be extended when the evidence presented to court is a model output rather than a direct reproduction of the underlying file. Digital forensic guidance treats documentation and preservation as integral to defensible examination (Kent et al., 2006).

For AI-assisted processing, an adequate record should identify the source file, relevant preprocessing, model or software version, operator, dates, configuration settings, and the output actually produced. Whether an enhancement increases sharpness, a transcription engine reconstructs noisy speech, or a generative model summarizes messages, the prosecution must be able to distinguish observed data from newly inferred or synthesized content. Otherwise, the court may mistake an algorithmic reconstruction for a feature that was visible or audible in the original record. This confusion is particularly problematic with deepfake detection, where the classifier itself may misidentify authentic media or overlook manipulated material (Chesney & Citron, 2019; Verdoliva, 2020).

Cloud-based tools add uncertainty because service providers may update models or alter processing pipelines after an initial examination. A later test using the same brand of software does not necessarily replicate the prior system. The legally material question is whether a qualified examiner can reconstruct the relevant operational conditions or explain why that is impossible. A strict demand for identical results in every probabilistic system would be unrealistic, but the absence of version information or meaningful testing records may severely weaken an asserted inference. Traceability must therefore attach not only to the electronic file but also to its pathway through computational transformations.

Technical Weaknesses: Validity, Accuracy, Bias, and Interpretation

A third result is the need to separate technological capability from courtroom proof. Model validity addresses whether the system is suitable for the proposed task; reliability concerns consistency under defined conditions; and accuracy concerns measured performance relative to a ground truth or evaluation criterion. Each needs to be contextualized. A vendor's reported recognition accuracy may come from high-quality benchmark images and tell the court little about a blurred nighttime video. Similarly, a percentage attached to a facial match may be a similarity score or ranking metric rather than the probability that the person is guilty, or even the probability of identity (Grother et al., 2019).

Error rates must be presented in terms relevant to the claim. A false-positive match can wrongly nominate an individual for further investigation, whereas a false negative can omit an appropriate candidate. A model may also operate with thresholds that change the balance between the two errors. An expert should explain the evaluation population, data quality, operating threshold, and whether the tested task matches the case-specific task. Accuracy cannot be interpreted sensibly without the conditions under which it was measured. The NIST framework provides a conceptual foundation for measuring risks and documenting limitations, although it is not itself an Indonesian rule of criminal evidence (Tabassi, 2023).

Bias is not confined to discriminatory intention. Uneven dataset coverage or performance under different environmental conditions can produce systematically different error patterns. Facial recognition, automated speech analysis, and predictive classification may reproduce deficiencies in collection practices or training data. Barocas and Selbst (2016) show how data-driven systems can generate disparate effects even without express discriminatory instructions. In a criminal trial, therefore, an allegation of algorithmic bias should lead to a focused inquiry into whether the identified limitation is relevant to the particular comparison or inference being offered, not to an assumption that every AI system is either inherently biased or inherently neutral.

Technical opacity creates a separate evidentiary weakness. Systems may be difficult to explain because their internal operations are complex or because their design is commercially protected. Neither circumstance justifies blind acceptance. Relevant information might include model documentation, validation studies, confidence-score interpretation, logs, independent testing, and the qualifications of the analyst. A demand for source code in every case would be disproportionate, while denial of all meaningful information can frustrate the right to challenge adverse evidence. The appropriate scope of disclosure should follow the materiality of the contested inferential claim and the defense's demonstrated need (Burrell, 2016; Rudin, 2019).

Comparative Result: Different Lessons from Loomis and Miles

State v. Loomis (2016) offers a caution against inflating machine-generated assessments into determinations of individual guilt or appropriate punishment. The case addressed Wisconsin's use of COMPAS scores during sentencing. The court allowed a bounded role for the assessment, with cautions about its limitations, rather than treating the proprietary score as an independent substitute for judicial reasoning. Because it concerns sentencing after conviction, it should not be cited as a direct rule governing the admission of AI-produced trial evidence. Its transferable lesson is narrower and stronger: algorithmic outputs must not displace individualized assessment and reasoned human responsibility.

State v. Miles (2026), by contrast, addresses facial-recognition discovery during an investigation into serious criminal charges. The New Jersey Supreme Court required the prosecution to disclose information identifying the technology used and showing how investigators used it, given the potential relevance to reliability, witness-identification challenges, investigative errors, and alternative suspects. It rejected a mechanical assumption that every proprietary item must be handed over, leaving source-code disclosure open to a later showing of particularized need. The decision shows that relevant procedural rights may arise even when the technology was presented as an investigative lead rather than as the central trial exhibit. This principle has analytical value for Indonesia but cannot be transplanted as a binding disclosure rule.

Taken together, the cases illustrate a differentiated response to technological evidence. Loomis stresses limits on the authority granted to a predictive assessment; Miles emphasizes targeted access to information necessary to challenge the investigation. Both reject the simplistic premise that machine output speaks for itself. Neither resolves Indonesian statutory interpretation or determines a uniform scientific threshold for admissibility. The comparative result supports designing a framework that is sensitive to procedural stage, the actual role of the model, and the importance of the disputed output to an individual's defense.

The Proposed Two-Layer Assessment Framework

The fourth and principal finding is a two-layer evaluation sequence. Its first layer is procedural: identify the precise evidentiary item, the legal authority through which it was obtained, its authenticity, integrity, and documented chain of custody. For an AI-produced derivative, the inquiry should also record the transformation from original data to output. When the underlying evidence fails the statutory conditions for authenticity or lawful acquisition, the consequences should be governed by Article 235, not repaired through a favorable technical performance score. This preserves the distinction between legal permission to use evidence and confidence in a computational conclusion.

The second layer examines technical reliability: task appropriateness, model identification, validation context, quality of input data, reported error rates, relevant bias, interpretability, reproducibility, and the possibility of independent review. Its intensity should vary with the evidentiary function. A model used to sort thousands of messages for an initial investigation does not automatically call for the same scrutiny as a facial-recognition identification that is offered to establish the defendant's presence at a crime scene. In either setting, however, a material output should be accompanied by enough information for a court to understand its limitations and for the defense to contest its significance.

The framework concludes with an integrated judicial appraisal rather than a mechanically automatic admissibility decision. The judge must ask what fact the output actually tends to establish, how strongly it does so, and whether independent evidence supports or contradicts it. A high similarity score may justify examining a possible identity, yet may not prove presence, participation, intention, or guilt. A reliable deepfake detector may show signs of manipulation, yet may not establish who manipulated a file. These are inferential limits, not simply technology defects. Separating them prevents the court from translating a computational label into an unsupported conclusion about statutory elements.

The framework should be understood as a proposal for implementation guidance. Article 235 supplies existing legal rules on authentication and lawful collection; the additional validity, explainability, and bias considerations are prescriptive standards proposed through this research. They are not represented as requirements already enumerated in the Code. This distinction matters for legal certainty and institutional feasibility: judicial reasoning can draw on scientific evidence now, while detailed administrative protocols, expert qualifications, preservation formats, and independent audit arrangements may require further regulation and resources.

DISCUSSION

From Statutory Admissibility to Responsible Evaluation of Inference

The proposed reconstruction responds to a recurrent category error: treating the legally authentic status of a data source as confirmation of every conclusion subsequently generated from it. Article 235 addresses important threshold problems, but its language should not be stretched into an assertion that authentication alone validates a model. Technical assessment does not compete with the authority of law; it supplies the factual basis for rationally applying legal standards to complex evidence. A court must first know which claim is being made. If the claim is that a CCTV file is an original recording, provenance and integrity are central. If the claim is that a facial-recognition output accurately identifies the defendant, the court must additionally examine what the system measured and what uncertainty remains.

A two-layer structure is also preferable to an undifferentiated demand that all AI evidence satisfy a single accuracy threshold. A universal minimum would be difficult to justify because tools have different functions and the costs of particular errors vary. A translation system used to search for relevant files, for example, is not equivalent to a biometric classifier presented as the basis for identifying a suspect. The same numerical performance may also have different significance depending on the size of the comparison database and the quality of the submitted input. What matters is whether the method is sufficiently supported for its proposed evidentiary purpose and whether the judge identifies limits on the inference (Grother et al., 2019; Tabassi, 2023).

Proportionate Disclosure as a Procedural Safeguard

Meaningful contestation requires access to information, but it need not entail unrestricted disclosure of sensitive personal records or a technology vendor's entire intellectual property. A proportionate response begins with a basic disclosure package covering the model's identity, version, relevant operation, input transformations, performance characteristics, and the output used. If that package leaves a material reliability question unresolved, the defense should be able to seek more focused testing, expert examination, or access subject to protective arrangements. The court can assess necessity in light of the case rather than presume that all proprietary materials are either always indispensable or always immune from scrutiny. This is consistent with the calibrated reasoning illustrated by State v. Miles (2026).

Importantly, the use of a model as an investigative lead should not automatically prevent scrutiny of its influence on later evidence. An early facial-recognition candidate list may shape which witnesses are shown photographs, which leads are pursued, and which persons are excluded. Later testimonial identifications might appear independent while having been influenced by an initial computational suggestion. The inquiry should therefore trace how an algorithmic lead informed subsequent investigative decision, while still distinguishing that investigatory role from the formal evidentiary status of an exhibit. Such an account protects the defendant's ability to challenge the development of a case without turning every exploratory computer search into a separate full evidentiary hearing.

This approach also requires careful handling of third-party privacy. A database used for facial recognition may include personal information about many people uninvolved in the offense. Broad disclosure of raw data could harm those individuals without materially improving the defense. A court may instead consider redacted records, expert summaries, restricted inspection, or narrowly tailored access if permitted by applicable law. The legal objective is effective challenge to a consequential inference, not indiscriminate circulation of sensitive information. Reconciling disclosure and protection is especially important when the underlying data include biometric identifiers governed by Indonesia's personal-data framework.

Expert Evidence and The Judge's Responsibility to Give Reasons

Expert assistance is indispensable but cannot be treated as a replacement for judicial reasoning. A digital-forensics examiner may competently explain file acquisition and hash verification without being qualified to interpret machine-learning validation statistics. Conversely, a model developer may understand algorithmic design but lack knowledge of whether police acquired a file through lawful procedures. Courts should match the expert's demonstrated competencies to the specific disputed question and require a clear separation between observations, assumptions, model outputs, and inferences. The quality of expert testimony depends as much on acknowledging limitations as on presenting a favorable conclusion (Garrett & Neufeld, 2009; Murphy, 2007).

Judges need not become programmers or statisticians. They do, however, need a disciplined set of questions: What exactly did the technology do? Was that task validated for data like the case material? What were the relevant false-match or other error rates? Did an operator change settings or edit inputs? Could another qualified examiner meaningfully review the result? Which proposition does the output establish, and which remains unproved? Answering these questions helps the court guard against automation bias, the tendency to give an apparently objective score more weight than its methodological foundation warrants. Reasons for judgment should articulate why a particular output was credited, discounted, or excluded rather than merely repeat an expert's conclusion (Roth, 2016).

An illustrative facial-recognition dispute clarifies the distinction. Assume police lawfully obtain a nighttime recording, preserve the original, and submit a frame to an AI system that returns a high-ranked candidate. The first layer asks whether collection was authorized, the recording is authentic, and processing is documented. The second asks about image quality, model version, database construction, score meaning, threshold choice, and comparable performance tests. Even if both inquiries produce favorable answers, an additional inference is required: whether the person was actually involved in the charged conduct. Presence, identity, participation, and intent are different legal propositions. Other evidence, including alternative explanations and alibi material, remains essential to the final judgment.

Institutional Implementation and Normative Restraint

Operational adoption should proceed through phased professional guidance rather than an assumption that every court can immediately audit complex software. Investigators could use standardized logs distinguishing original evidence, preprocessing, system identity, and output. Prosecutors could examine whether a technical report identifies sources of error and is capable of adversarial testing. Defense lawyers could use structured applications for relevant disclosure, and courts could maintain a checklist of the legal and inferential questions that require answers. Cross-disciplinary training would help each institution recognize the boundaries of its expertise without pretending that one professional can independently validate every form of AI (Sumardiana et al., 2024).

The proposed framework is deliberately technology-neutral. Naming a particular vendor or currently popular model as the measure of reliability would quickly become obsolete. Function-based criteria travel better across facial recognition, speaker comparison, deepfake assessment, communication clustering, and generative summaries. Nevertheless, technology-neutrality does not mean that every application is identical. A generative summary should be verified against source records because fabricated detail is a plausible risk; a facial-recognition match calls for context-specific identity testing; and a deepfake classification invites inquiry into the detector's validation and the file's provenance. The common legal architecture is stable even though the technical questions must be adapted to the tool and task.

There are also principled limits to reconstruction. The framework cannot alter legislatively prescribed exclusionary consequences, shift the burden of proving a criminal charge to the accused, or authorize judges to compel information outside lawful procedure. Neither NIST guidance nor a foreign judgment has direct normative force as an Indonesian evidentiary provision. Their function in this study is to illuminate questions that existing legal doctrine should confront and to guide future procedural detail. Failure to distinguish descriptive law from proposed policy could undermine the very legal certainty the framework seeks to strengthen. Accordingly, any regulatory implementation should explicitly state legal authority, scope, safeguards, and remedies for noncompliance.

The significance of the model ultimately lies in strengthening, rather than displacing, human adjudication. AI can identify patterns that investigators would otherwise miss and assist the handling of voluminous digital material. Yet a judicial decision remains an exercise in evaluating disputed facts under legal standards and explaining why the resulting conclusion is justified. The two-layer approach insists that an electronic source be legally usable, an algorithmic result be sufficiently reliable for its claimed purpose, and a final inference be independently supported. This integrated discipline offers a practical conceptual bridge between Indonesia's new Criminal Procedure Code and the scientific uncertainties of contemporary criminal evidence.

CONCLUSIONS AND RECOMMENDATIONS

Indonesia's Law Number 20 of 2025 has strengthened the formal position of electronic evidence through explicit recognition, authentication requirements, judicial scrutiny of lawful acquisition, and exclusionary consequences. Nevertheless, these provisions do not by themselves supply detailed standards for assessing AI-generated classifications, identifications, or predictions. The central problem is an evaluative gap between the legal status of an electronic source and the scientific reliability of the conclusions produced from it. Conflating those matters risks assigning unwarranted weight to algorithmic outputs and weakening meaningful defense challenges.

This article proposes a two-layer assessment framework. The procedural layer examines lawful acquisition, authenticity, integrity, custody, and traceable data transformation. The technical layer examines task validity, contextual performance, error rates, bias, interpretability, and the opportunity for independent scrutiny. A separate judicial appraisal then connects the limited factual proposition established by the output to the elements of the offense and the other evidence. Targeted disclosure, competent expert assistance, proportionate review, and reasoned judicial decisions are necessary safeguards. Indonesian institutions should develop practical documentation and evaluation guidance while preserving statutory limits and defendants' procedural rights

ADVANCED RESEARCH

This study is normative and has not tested the framework against an empirical sample of Indonesian prosecutions or measured the accuracy of particular deployed systems. Future research should examine actual judicial records, defense access to technical material, and investigative practices concerning facial recognition, audio processing, and synthetic-media detection. Multidisciplinary pilot studies could assess which disclosure elements and validation reports most improve judicial understanding without compromising privacy or legitimate confidentiality. Further comparative research should examine how varied legal systems distinguish investigative leads from proof offered at trial and how they allocate responsibility for preserving and explaining algorithmic transformations. Such work would permit refinement of the framework into tested, context-sensitive procedural guidance.

ACKNOWLEDGMENT

The author acknowledges the academic environment of the Faculty of Law, Universitas Atma Jaya Yogyakarta, which supports inquiry into contemporary criminal-procedural challenges. No external funding is asserted in this draft.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732. https://doi.org/10.15779/Z38BG31.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1). https://doi.org/10.1177/2053951715622512.

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107(6), 1753–1820. https://doi.org/10.15779/Z38RV0D15J.

Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. Washington Law Review, 89(1), 1–33.

Garrett, B. L., & Neufeld, P. J. (2009). Invalid forensic science testimony and wrongful convictions. Virginia Law Review, 95(1), 1–97.

Gless, S. (2020). AI in the courtroom: A comparative analysis of machine evidence in criminal trials. Georgetown Journal of International Law, 51(2), 195–253.

Gless, S., Lederer, F. I., & Weigend, T. (2024). AI-based evidence in criminal trials? Tulsa Law Review, 59(1), 1–37.

Grother, P., Ngan, M., & Hanaoka, K. (2019). Face recognition vendor test (FRVT) part 3: Demographic effects (NIST Interagency Report 8280). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.8280.

Kent, K., Chevalier, S., Grance, T., & Dang, H. (2006). Guide to integrating forensic techniques into incident response (NIST Special Publication 800-86). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-86.

Murphy, E. (2007). The new forensics: Criminal justice, false certainty, and the second generation of scientific evidence. California Law Review, 95(3), 721–797. https://doi.org/10.15779/Z38R404.

Ramiyanto. (2017). Bukti elektronik sebagai alat bukti yang sah dalam hukum acara pidana. Jurnal Hukum dan Peradilan, 6(3), 463–484. https://doi.org/10.25216/jhp.6.3.2017.463-484.

Republic of Indonesia. (1945). Constitution of the Republic of Indonesia of 1945, Articles 28D (1) and 28G (1).

Republic of Indonesia. (2008). Law Number 11 of 2008 on Electronic Information and Transactions (as subsequently amended by Law Number 19 of 2016 and Law Number 1 of 2024).

Republic of Indonesia. (2019). Government Regulation Number 71 of 2019 on the Operation of Electronic Systems and Transactions.

Republic of Indonesia. (2022). Law Number 27 of 2022 on Personal Data Protection.

Republic of Indonesia. (2025). Law Number 20 of 2025 on the Criminal Procedure Code.

Roth, A. L. (2016). Trial by machine. Georgetown Law Journal, 104(5), 1245–1305.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x.

State v. Loomis, 2016 WI 68, 371 Wis. 2d 235, 881 N.W.2d 749 (Wisconsin Supreme Court, 2016).

State v. Miles, A-41-24 (090275) (Supreme Court of New Jersey, June 24, 2026). https://www.njcourts.gov/cases/a-41-24.

Sumardiana, B., Pujiyono, P., Cahyaningtyas, I., & Wulandari, C. (2024). Evaluation of electronic evidence in criminal justice in the era of advanced artificial intelligence technology. Indonesian Journal of Criminal Law Studies, 9(2), 309–332. https://doi.org/10.15294/ijcls.v9i2.36634.

Tabassi, E. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1.

Verdoliva, L. (2020). Media forensics and deepfakes: An overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910–932. https://doi.org/10.1109/JSTSP.2020.3002101.

Wachjoe P., S. (2016). Penggunaan informasi elektronik dan dokumen elektronik sebagai alat bukti persidangan. Jurnal Hukum dan Peradilan, 5(1), 1–18. https://doi.org/10.25216/jhp.5.1.2016.1-18.