

Reconstructing Validity and Reliability Standards for AI-Based Evidence in Indonesian Criminal Proceedings

Gregorius Widiartana^{1*}, Vincentius Patria Setyawan²

^{1,2}Universitas Atma Jaya Yogyakarta, Indonesia

*Corresponding author: gwidiartana0@gmail.com

Received: August 10, 2026 | Revised: September 5, 2026 | Accepted: September 11, 2026 | Published: September 30, 2026

Article DOI: <https://doi.org/10.60153/ijolares.v4i2.372>

ABSTRACT

The use of artificial intelligence in criminal investigations and court proceedings creates new challenges for Indonesian evidence law. AI can be used for facial recognition, data analysis, voice identification, and detecting fake media. However, Indonesian criminal procedure does not yet provide clear rules on whether AI-based evidence is reliable, transparent, and acceptable in court. This study aims to develop a normative framework for evaluating the admissibility and reliability of artificial intelligence-based evidence in Indonesian criminal proceedings by integrating evidentiary principles, algorithmic accountability, and defendants' procedural rights. The research employs a normative legal method using statutory, conceptual, and comparative approaches, examining Indonesian evidentiary rules alongside selected foreign legal frameworks governing the admissibility and scrutiny of artificial intelligence-based evidence. Primary legal materials include Indonesian criminal procedure legislation and regulations concerning electronic information, while secondary materials consist of legal scholarship on digital evidence, algorithmic accountability, and fair trial guarantees. The study finds that classifying evidence as electronic evidence is not enough to establish its admissibility. AI-based evidence may raise additional concerns about its source, reliability, transparency, and the ability of the defence to challenge how it was produced and interpreted. Its admissibility should depend on how the evidence was obtained, its integrity and accuracy, the transparency of the methods used, expert verification, and the defence's ability to examine and challenge it. The study concludes that Indonesia needs a clear legal framework for AI-based evidence to ensure that AI results are properly examined and that the use of technology does not undermine due process.

Keywords: artificial intelligence-based evidence; criminal proceedings; evidentiary standards

1. INTRODUCTION

The rapid development of artificial intelligence (AI) has changed the manner in which facts are discovered, processed, and presented in criminal proceedings. Law-enforcement institutions and forensic experts may use AI systems for facial and voice recognition, the automated examination of large datasets, image enhancement, pattern identification, cyber-forensic analysis, and the detection of manipulated audio-visual content. At the same time, generative AI can produce synthetic images, voices, videos, and written materials that closely resemble authentic records. These developments create distinct evidentiary challenges because AI may either assist in processing pre-existing evidence or materially generate, transform, or interpret the evidentiary content itself, thereby requiring different standards for assessing authenticity, reliability,

and probative value. AI may assist in identifying and interpreting evidence, but it may also generate or transform the material submitted as evidence. Consequently, the central legal issue is no longer limited to whether electronic information may be admitted in court. It also concerns whether an AI-based output accurately represents the underlying data, whether the method producing the output is scientifically dependable, and whether the accused has a meaningful opportunity to examine and challenge it.

For purposes of this study, 'AI-based evidence' refers to evidentiary material whose generation, transformation, identification, classification, reconstruction, or interpretation materially depends on an AI system in a manner that may affect its authenticity, reliability, or probative value. This definition distinguishes AI-based evidence from ordinary electronic evidence

that merely exists in digital form. A conventional digital photograph, for example, may primarily raise questions regarding its origin, integrity, and chain of custody. By contrast, an AI-assisted facial-recognition match or an algorithmically reconstructed recording additionally raises questions concerning training data, model design, performance limitations, error rates, deployment conditions, and the relationship between the system's output and the fact to be proven. Machine-learning systems must therefore be examined as socio-technical processes rather than as neutral devices because their results depend on choices made during problem definition, data collection, data preparation, model selection, training, and deployment (Lehr & Ohm, 2017).

The distinction is particularly important in criminal adjudication, where an evidentiary error may affect personal liberty and the legitimacy of a conviction. Machine-produced assertions may therefore raise two distinct legal questions: whether the output satisfies the applicable conditions for admission and, once admitted, what probative weight should be assigned to it in light of its methodological reliability, error rate, and evidentiary limitations. A machine output may appear objective because it is expressed numerically or produced automatically, although its accuracy remains dependent on data quality, methodological assumptions, software configuration, and human intervention. (Biedermann & Lau, 2024) further demonstrate that reliance on machine evidence cannot be reduced to aggregate accuracy alone because the probative meaning of computational output must be assessed in relation to the specific decision and the risks of an incorrect outcome. Moreover, proprietary systems may prevent defendants and independent experts from examining source code, validation studies, or operational parameters. (Wexler, 2018) demonstrates that trade-secret claims in criminal proceedings can obstruct access to information necessary to test forensic technologies, thereby creating tension between commercial confidentiality and the accused's rights to liberty and a fair process. The reliability of AI-based evidence thus cannot be presumed

solely because the output was generated by sophisticated technology or presented through expert testimony.

Indonesia has recently strengthened the formal legal basis for electronic evidence. Law Number 20 of 2025 concerning the Criminal Procedure Code, which has been effective since 2 January 2026, expressly recognises electronic evidence as one of the lawful forms of evidence in Article 235(1)(f). Article 242 further provides that electronic evidence includes electronic information, electronic documents, and electronic systems related to a criminal offence. In addition, Article 235(5) excludes evidence declared unauthentic or obtained unlawfully (Republic of Indonesia, 2025). This framework operates together with Article 5 of Law Number 11 of 2008 concerning Electronic Information and Transactions, as last amended by Law Number 1 of 2024, which recognises electronic information, electronic documents, and their printouts as lawful evidence (Republic of Indonesia, 2024). These provisions represent an important shift from the former Criminal Procedure Code, under which electronic evidence was mainly recognised through sectoral legislation. Although Articles 235 and 242 establish the legal status of electronic evidence, they do not expressly determine how courts should assess the evidentiary reliability of AI-generated or AI-processed outputs, particularly where the probative result depends on model performance, training data, system configuration, or algorithmic processing that cannot be evaluated through conventional authenticity requirements alone.

The present framework primarily addresses the legal status and general scope of electronic evidence. It does not expressly regulate the minimum disclosure required for an AI system, the validation of its model and dataset, acceptable error rates, reproducibility, explainability, independent auditing, version control, or the procedure through which the defence may contest an algorithmic output. Authenticity and lawful acquisition are indispensable, but they do not fully answer

whether a system is reliable for the particular purpose and population involved in a criminal case. An authentic report may faithfully reproduce the output generated by a system while the system itself applies an unsuitable model, relies on incomplete data, or produces an unacceptable rate of false identification. Conversely, a technically reliable tool may still yield legally invalid evidence when the underlying data were acquired unlawfully or when the chain of custody was broken. Validity and reliability are therefore related but distinct dimensions that must be tested separately and cumulatively. This distinction is consistent with the broader view that trustworthy AI requires the integration of legal requirements and technical assessment rather than reliance on technical performance in isolation (Kattnig et al., 2024).

Indonesian scholarship has begun to address this technological transition. (Sumardiana et al., 2024) examine the position of electronic evidence in the era of advanced AI and emphasise the need for ethical frameworks, institutional capacity, and continuous legal adaptation. (Permadi et al., 2026) focus on deepfake-based evidence in fraud cases and propose a reliability framework based on digital-forensic verification, metadata validation, and chain of custody. More recently, (Antoro et al., 2026) compare Indonesia, the United States, and Japan and propose five parameters for the admissibility of AI-related electronic evidence: authenticity and integrity, process traceability, model performance, identity verification, and accountability. These studies make important contributions by demonstrating that the general recognition of electronic evidence is insufficient for AI-related materials.

Existing scholarship addresses important aspects of AI-related evidence, but its analytical focus remains fragmented. Studies concerning electronic evidence generally emphasise authenticity and integrity, while research on AI-generated or manipulated evidence tends to focus on technical reliability or specific applications such as deepfakes. Other approaches address admissibility without fully examining the

procedural rights necessary for meaningful defence challenges. The literature therefore lacks an integrated framework that connects legal validity, technical reliability, and procedural contestability within the Indonesian criminal procedure system, even though AI may perform materially different evidentiary functions: generating content, detecting manipulation, identifying a person, classifying data, reconstructing an event, or producing an expert-like inference. Each function creates different risks and therefore requires a context-sensitive standard. Second, prior studies tend to combine admissibility, authenticity, and reliability without clearly separating their legal and scientific consequences. Third, the literature has not sufficiently integrated technical assessment with procedural contestability under the architecture of the new Indonesian Criminal Procedure Code. A reliable evidentiary framework must address not only how judges assess an AI system but also what information must be disclosed, who bears the initial burden of establishing reliability, how an independent examination may be conducted, and what consequence follows when meaningful scrutiny is impossible.

This article addresses those gaps by reconstructing a layered standard for the validity and reliability of AI-based evidence in Indonesian criminal proceedings. The proposed model consists of three cumulative levels. The first is legal validity, covering lawful acquisition, relevance, authenticity, integrity, and an uninterrupted chain of custody. The second is scientific and technical reliability, covering the suitability of the AI system for its stated forensic purpose, data quality, documented validation, known limitations and error rates, reproducibility, version identification, and independent auditability. The third is procedural contestability, covering disclosure of materially relevant information, expert verification, equality of arms, reasoned judicial assessment, and the defendant's effective right to challenge the evidence. This layered structure avoids two opposite errors: automatically rejecting AI-based evidence because it is technologically complex or

automatically accepting it because it appears scientific and efficient.

The proposed distinction also determines the legal consequences of a failure at each level. Evidence that fails the legal-validity test should be excluded when it was unlawfully obtained, cannot be authenticated, or has lost its integrity. Evidence that satisfies formal validity but fails the technical-reliability test should not be treated as sufficient proof of the fact asserted by the AI output. Meanwhile, evidence whose operation cannot be meaningfully examined by the defendant raises a procedural-fairness problem, even when the prosecution claims that the system has been independently validated. The judicial assessment must therefore move beyond a binary question of admissibility and consider whether the output possesses adequate probative value, whether it requires corroboration, and whether its limitations have been transparently communicated in court.

Accordingly, this study asks two related questions: how are the validity and reliability of AI-based evidence presently accommodated within Indonesian criminal procedure law, and how should the applicable standards be reconstructed to protect evidentiary accuracy and due process? The study aims to formulate a judicially applicable framework that connects statutory legality, technical dependability, and procedural fairness. Its novelty lies not merely in adding another list of technical indicators but in organising those indicators according to their distinct legal functions and consequences. Unlike previous models that primarily formulate general admissibility parameters or focus specifically on deepfake evidence, this study differentiates the legal validity of the evidence, the scientific reliability of the AI system, and the procedural conditions necessary to contest its output.

Through this reconstruction, AI-based evidence may be used to support the search for material truth without transforming opaque algorithmic outputs into uncontestable assertions. The proposed framework is expected to guide judges, prosecutors, investigators, defence counsel, and forensic experts in determining the

admissibility and probative value of AI-based evidence. It may also provide a normative basis for future implementing regulations, Supreme Court guidelines, digital-forensic protocols, and procedures for the disclosure and independent examination of AI systems used in Indonesian criminal proceedings.

2. METHOD

This study employed normative legal research to assess whether Indonesian criminal procedure provides sufficient legal standards for the admissibility, authentication, reliability, and contestability of artificial intelligence-based evidence. Normative legal research was selected because the principal research problems concern the interpretation, consistency, and development of legal norms rather than the measurement of social behaviour or institutional practices. This research design enabled the study to identify the applicable evidentiary rules, evaluate their capacity to respond to the distinctive characteristics of artificial intelligence systems, and formulate a prescriptive legal framework for judicial assessment. A normative approach was adopted because the study seeks to interpret existing evidentiary norms, identify regulatory gaps, and formulate prescriptive standards for the judicial assessment of AI-based evidence (Marzuki, 2017; Soekanto & Mamudji, 2015).

The research applied statutory, conceptual, case, and comparative approaches. The statutory analysis examined provisions governing the admissibility, acquisition, authenticity, integrity, and exclusion of electronic evidence, with particular attention to their applicability to AI-generated and AI-processed evidentiary outputs. The analysis focused on Law Number 20 of 2025 concerning the Criminal Procedure Code as the relevant statutory framework in force during the study period, particularly its provisions governing lawful and electronic evidence, authenticity, and exclusion of unlawfully obtained evidence, and Law Number 11 of 2008 concerning Electronic Information and Transactions, as most recently amended by Law Number 1 of 2024. Relevant implementing

regulations and official legal documents concerning electronic systems, personal data protection, digital forensics, and criminal adjudication were also examined. This approach was intended to identify normative gaps between the general recognition of electronic evidence and the absence of specific standards for evidence generated, processed, or materially influenced by artificial intelligence systems.

The conceptual analysis grouped the relevant concepts into three dimensions: legal validity and admissibility, technical reliability and integrity, and procedural contestability and defence rights. The conceptual analysis distinguished authenticity, which concerns the provenance and integrity of the evidentiary material, from reliability, which concerns the validity and performance of the system producing the output, and from probative value, which concerns the weight assigned to the evidence by the court. The approach also examined the relationship between technological reliability and fundamental principles of criminal procedure, including due process of law, the presumption of innocence, equality of arms, the right to defence, and the judicial obligation to provide reasoned decisions. Legal concepts and doctrines were analysed to construct a layered framework that differentiates legal validity, technical reliability, and procedural fairness.

The case approach was used to examine judicial reasoning in selected foreign cases involving algorithmic or software-assisted evidence. The cases were purposively selected because they provide distinct judicial approaches to algorithmic opacity, disclosure, reliability, and the defendant's ability to challenge computational evidence, limitations on disclosure, the reliability of computational outputs, and the defendant's opportunity to challenge technology-based evidence. The cases were not treated as binding authorities for Indonesian courts but as analytical materials for identifying legal principles relevant to transparency, verification, and procedural contestability.

The comparative approach examined selected legal standards from the United States and the European Union. The United States and European Union were selected because they provide complementary regulatory responses to the evidentiary challenges examined: judicial gatekeeping and scientific reliability in the former, and risk-based governance, transparency, and human oversight in the latter. The European Union was selected because Regulation (EU) 2024/1689, or the Artificial Intelligence Act, establishes a risk-based regulatory framework that includes requirements concerning risk management, data governance, technical documentation, record-keeping, transparency, human oversight, accuracy, robustness, and cybersecurity for high-risk artificial intelligence systems. The functional comparison examined four dimensions: admissibility, reliability assessment, transparency and disclosure, and procedural safeguards, which were subsequently assessed for compatibility with Indonesian criminal procedure. It examined how each system responds to similar problems of reliability, opacity, accountability, and fundamental-rights protection and then assessed which principles could be adapted to the structure of Indonesian criminal procedure.

The legal materials consisted of primary and secondary legal materials. Primary legal materials included Indonesian legislation and regulations, selected foreign legislation, official regulatory documents, and judicial decisions concerning electronic, scientific, digital, or algorithmic evidence. Secondary legal materials consisted of peer-reviewed journal articles, scholarly books, legal commentaries, and interdisciplinary studies on criminal evidence, artificial intelligence, digital forensics, algorithmic accountability, and fair-trial rights. Tertiary materials, including legal dictionaries and institutional glossaries, were used only to clarify specialised legal or technological terminology.

The document study covered relevant legal databases, official institutional sources, and academic databases using predefined search

terms and a specified publication period, supplemented by foundational authorities where necessary. Legislation and judicial decisions were identified through official legal databases and institutional websites, while scholarly literature was identified through searches using combinations of the terms “artificial intelligence-based evidence,” “algorithmic evidence,” “machine evidence,” “electronic evidence,” “digital forensics,” “algorithmic transparency,” “evidentiary reliability,” “criminal proceedings,” and “due process.” Materials were screened against predefined inclusion criteria and retained when they directly addressed AI-related evidence, evidentiary reliability, procedural contestability, or regulatory standards relevant to the research questions. Recent publications were prioritised to reflect technological and legal developments, while earlier sources were retained when they provided foundational doctrines or influential analytical frameworks. The collected materials were classified according to the source of law, jurisdiction, type of artificial intelligence application, evidentiary issue, and legal principle addressed.

The study employed qualitative doctrinal analysis conducted in four stages. First, statutory interpretation was used to determine the scope and meaning of Indonesian provisions governing electronic evidence. Second, conceptual analysis distinguished the legal, technical, and procedural dimensions of artificial intelligence-based evidence. Third, case and comparative analyses identified common principles and alternative regulatory responses from the selected jurisdictions. Fourth, the identified legal principles and comparative findings were synthesised into three analytical dimensions: legal validity, technical reliability, and procedural contestability based on their distinct legal functions and consequences for the admissibility and probative value of AI-based evidence. The proposed reconstruction was evaluated according to its consistency with Indonesian criminal procedure, its capacity to protect the defendant’s rights, and its practical applicability in judicial assessment. This

analytical sequence connected the selected legal materials directly to the research questions and enabled the formulation of a systematic legal solution (Fajar & Achmad, 2010). Because this study did not involve human participants, interviews, surveys, experiments, or the collection of personal data, empirical sampling and participant-related ethical approval were not applicable.

3. RESULTS AND DISCUSSION

3.1 The Normative Position of Artificial Intelligence-Based Evidence in Indonesian Criminal Procedure

The analysis produces three principal findings. First, Indonesian criminal procedure already provides a legal entry point for the admission of artificial intelligence-based evidence through the recognition of electronic evidence. Second, the existing framework predominantly regulates the formal validity of evidence, particularly its authenticity and lawful acquisition, but does not establish an adequate standard for examining the scientific and technical reliability of an artificial intelligence system. Third, this normative gap can be addressed without necessarily creating an entirely new category of evidence. What is required is a special judicial assessment framework that operates within the existing categories of electronic evidence, expert evidence, documentary evidence, and physical evidence.

Article 235(1) of Law Number 20 of 2025 concerning the Criminal Procedure Code recognises electronic evidence as one of the lawful forms of evidence. The provision also recognises physical evidence, judicial observation, and other materials that may be used for evidentiary purposes, provided that they were not unlawfully obtained. Article 235(3) further requires evidence to be authenticable and lawfully obtained, while Article 235(4) authorises judges to assess its authenticity and legality. Under Article 235(5), evidence declared unauthentic or unlawfully obtained cannot be used at trial and has no probative force. Article 242 defines electronic evidence broadly as electronic information, electronic documents, or

electronic systems related to a criminal offence (Republic of Indonesia, 2025). These provisions establish authenticity and lawful acquisition as mandatory conditions for the use of electronic evidence.

This framework must be read together with Law Number 11 of 2008 concerning Electronic Information and Transactions, as last amended by Law Number 1 of 2024. The Electronic Information and Transactions Law recognises electronic information, electronic documents, and their printouts as lawful evidence. Government Regulation Number 71 of 2019 additionally requires electronic-system providers to operate electronic systems reliably, securely, and responsibly (Republic of Indonesia, 2019, 2024). These instruments demonstrate that Indonesian law already acknowledges reliability, security, integrity, accessibility, and accountability as important attributes of electronic systems. Nevertheless, those requirements primarily govern electronic-system operation and the general legal status of electronic information. They do not provide a detailed courtroom test for determining whether an AI-generated inference is sufficiently reliable to establish a disputed fact in a criminal case.

The central limitation lies in the distinction between the evidentiary object and the process through which its meaning is produced. Conventional electronic evidence usually consists of communications, recordings, transaction data, documents, metadata, or other information stored or transmitted electronically. Its examination generally focuses on whether the material originated from the claimed source, remained unaltered, and was obtained through lawful procedures. AI-based evidence adds another analytical layer because its evidentiary significance may be generated through a computational process. A facial-recognition result, for example, does not merely reproduce an image. It produces an inference concerning identity by comparing biometric features through a model, threshold, database, and matching process. Similarly, a deepfake detector does not merely display a recording but generates a

probabilistic or classificatory judgment concerning whether the recording has been synthetically created or manipulated.

AI-based evidence can therefore appear in at least four forms. The first is AI-generated evidentiary material, such as synthetic audio, images, or text allegedly used in committing an offence. The second is AI-processed evidence, where AI enhances, restores, translates, filters, or reconstructs existing material. The third is AI-derived inferential evidence, including facial recognition, speaker identification, behavioural classification, pattern detection, or predictive analysis. The fourth is AI-authentication evidence, where a system is used to determine whether digital content is authentic, manipulated, or synthetically generated. Each form involves different risks. An AI-generated fraudulent recording principally raises provenance and attribution issues, whereas an AI-based identification raises model-performance, database, threshold, and false-positive concerns. A single undifferentiated admissibility test would consequently be insufficient.

Existing Indonesian studies have correctly identified authenticity, traceability, model performance, chain of custody, and accountability as important considerations. (Sumardiana et al., 2024) emphasise the need for institutional capacity and ethical adaptation, (Permadi et al., 2026) propose forensic verification and metadata validation for deepfake evidence, and (Antoro et al., 2026) identify authenticity, integrity, traceability, performance, identity verification, and accountability as relevant admissibility parameters. The present study develops those contributions by separating three legally distinct questions: whether the evidence is legally valid, whether the AI system is scientifically and technically reliable, and whether its output can be meaningfully challenged in an adversarial criminal process.

Legal validity concerns whether the evidence may lawfully enter the adjudicative process. Technical reliability concerns whether the system performs accurately and dependably for the specific evidentiary purpose. Procedural

contestability concerns whether the accused has sufficient information and opportunity to question the system, its operator, its underlying data, and the interpretation of its output. Evidence may satisfy one dimension while failing another. A forensic report may be authentically signed and lawfully obtained but remain scientifically unreliable because the model was not validated for the relevant population. Conversely, an accurate system may produce legally invalid evidence when its input data were obtained through unlawful interception or when the integrity of the source material cannot be established.

Accordingly, AI-based evidence should not automatically be treated as a new and independent category of evidence. Its legal classification should follow its function in the proceedings. The original electronic file may constitute electronic evidence; the device or storage medium may constitute physical evidence; the system-generated report may constitute a document; and the explanation provided by a forensic specialist may constitute expert evidence. However, the involvement of AI should activate enhanced reliability and disclosure requirements. This functional approach prevents evidentiary classification from becoming technologically obsolete whenever a new model or application emerges.

3.2 The Inadequacy of Conventional Authenticity Tests

The second finding is that conventional authenticity and chain-of-custody requirements remain necessary but are insufficient for AI-based evidence. Authenticity ordinarily answers whether an item is what the proponent claims it to be. In AI-based evidence, authenticity may establish that a report is genuinely the report produced by a particular system. It does not establish that the system's conclusion is accurate. A facial-recognition report may be authentic as a document while the identification contained in it is erroneous. Similarly, a deepfake-detection report may accurately record the detector's score but the detector may perform poorly on

compressed Indonesian-language content or on material generated by a model absent from its training data.

AI systems are socio-technical systems whose outputs depend on the interaction between data, model design, human choices, deployment environments, and institutional procedures (Lehr & Ohm, 2017). Their performance cannot be inferred merely from the sophistication of their architecture or the reputation of their developer. The National Institute of Standards and Technology identifies validity and reliability as foundational characteristics of trustworthy AI and places them alongside safety, transparency, explainability, privacy, and fairness. It further stresses that accuracy measurements should be based on realistic and representative test sets, distinguish false-positive and false-negative rates, and reflect the intended conditions of use (Tabassi, 2023).

This context-specific requirement is particularly significant in criminal proceedings. An overall accuracy figure may conceal the kinds of error most relevant to the case. A system with high aggregate accuracy may still produce an unacceptable false-positive rate, resulting in an innocent person being identified as a suspect. Performance may also vary across datasets, demographic groups, image quality, language, accents, recording devices, environmental noise, and the size of the comparison database. Testing by the National Institute of Standards and Technology found substantial differences among facial-recognition algorithms and significant demographic variations in false-positive rates. The results also demonstrated that different systems should not be treated as possessing an equivalent level of reliability merely because they perform the same general function (Grother et al., 2019).

The same concern applies to AI systems used to detect synthetic content. Detection performance may decline when content has been resized, compressed, transmitted through social media, exposed to background noise, partially edited, or generated by a model that was not represented in the detector's training data.

Performance also differs across images, video, audio, and text. The National Institute of Standards and Technology notes that the effectiveness of many synthetic-content detection techniques remains incompletely evaluated and that detector performance depends heavily on training data, evaluation datasets, post-processing, and context of use (Chandra et al., 2024). Consequently, an AI detector's output should not be presented as a categorical statement that content is "real" or "fake" without communicating the score, decision threshold, uncertainty, error rates, and conditions under which the detector was evaluated.

The chain of custody must therefore be expanded into a computational chain of custody. Traditional chain-of-custody documentation records who collected, possessed, transferred, stored, and examined an evidentiary item. Digital-evidence guidance similarly emphasises the documentation, preservation, packaging, transportation, and forensic examination of electronic evidence (National Institute of Justice, 2008). For AI-based evidence, the record must additionally identify what operations were performed on the data and by which system.

A computational chain of custody should include the original input data, cryptographic hash values, time and method of acquisition, system and model name, model version, software configuration, parameters and thresholds, preprocessing operations, prompts where relevant, output files, audit logs, user identity, dates of analysis, and every subsequent transformation. Where an image was enhanced before being processed by a recognition model, both the original and enhanced versions must be preserved. Where generative AI was used to reconstruct a missing portion of a recording, the reconstructed portion must be clearly distinguished from the original material. A failure to separate generated reconstruction from observed fact could cause an algorithmic hypothesis to be misrepresented as historical reality.

Another risk is automation bias. Judges, investigators, and experts may assign excessive

weight to computational outputs because they appear numerical, neutral, or scientific. Machine-generated conclusions can nevertheless embed assumptions, data limitations, and human decisions that are invisible in the final report (Roth, 2017). A percentage, confidence score, or likelihood ratio must not be treated as self-explanatory. The court must understand what population, hypothesis, comparison method, and error structure the figure represents. Reliability therefore concerns not only correct software operation but also the validity of the inference drawn from the output.

3.3 Comparative Lessons for Indonesian Criminal Evidence

Comparative analysis demonstrates that transparency does not require the complete rejection of proprietary systems, but it does require safeguards against uncritical judicial reliance. In *State v. Loomis* (2016), the Wisconsin Supreme Court allowed consideration of a proprietary algorithmic risk assessment at sentencing, but imposed significant limitations. The court stressed that the assessment could not determine whether the defendant should be imprisoned, could not determine sentence severity, and could not constitute the determinative basis of the decision. It also required courts to be informed of limitations relating to proprietary design, population validation, demographic concerns, and the need for continued monitoring and recalibration. The decision illustrates that an AI output may be considered only within its validated purpose and should be accompanied by explicit warnings regarding its limitations.

However, *Loomis* should not be interpreted as authorising opaque systems to prove criminal guilt. The algorithm in that case was used at sentencing and as one factor among several independently established considerations. The risks are more serious when an AI output is introduced to prove identity, conduct, causation, intent, or another element of an offence. In such circumstances, the prosecution's burden and the presumption of innocence require stronger reliability and disclosure safeguards. An

identification generated by a black-box model cannot be equated with a risk-management recommendation used only to supplement a judge's evaluation.

A stronger approach to disclosure appears in *State v. Pickett* (2021). The New Jersey court held that when the prosecution relied on expert testimony generated through novel proprietary probabilistic-genotyping software, the defendant was entitled to discovery of the source code and related materials. The decision recognises that intellectual-property claims cannot eliminate the defendant's ability to test the mechanism underlying prosecution evidence. Confidential information may be protected through judicial orders, restricted expert access, or other procedural safeguards, but commercial secrecy cannot function as an absolute barrier to a fair defence.

The lesson for Indonesia is not that source code must be disclosed in every case. For some systems, disclosure of validation studies, documentation, model cards, error rates, input data, logs, and operational parameters may be sufficient. Source-code examination becomes necessary when the operation of the code is materially disputed, when documented testing cannot explain an anomalous output, when known software errors may affect the case, or when no less intrusive means enables effective verification. The court should balance commercial confidentiality against the seriousness of the charge, the evidentiary importance of the output, and the availability of alternative examination methods. Where meaningful verification remains impossible, the evidentiary consequence should fall on the party seeking to rely on the opaque system rather than on the accused.

The European Union Artificial Intelligence Act provides another relevant model, even though it is principally a regulatory instrument rather than a code of criminal evidence. The Act establishes requirements for high-risk AI systems concerning risk management, data governance, technical documentation, record-keeping, transparency, human oversight, accuracy,

robustness, and cybersecurity. AI systems used for certain law-enforcement purposes are treated as high risk because their use may significantly affect fundamental rights. These requirements demonstrate that reliable AI cannot be evaluated solely from the final output; governance and documentation throughout the system's lifecycle are equally important (European Parliament & Council, 2024).

These comparative models support three principles. First, AI must be used only for a purpose covered by its validation. Second, its limitations must be disclosed and judicially considered. Third, defendants must receive sufficient access to challenge outputs that materially affect their liberty. These principles can be adapted to Indonesia without transplanting foreign evidentiary doctrines wholesale.

3.3 Reconstruction of a Layered Standard

Based on the normative and comparative findings, this study proposes a three-layer standard comprising legal validity, scientific and technical reliability, and procedural contestability. The three layers are cumulative. Evidence should not be used to establish guilt merely because it satisfies one or two of them.

1) Legal Validity

The first layer examines whether the evidence satisfies the formal and procedural requirements of Indonesian law. The proponent must demonstrate relevance to the alleged offence, lawful acquisition, authenticity, integrity, and an adequately documented chain of custody. For AI-based evidence, authentication must cover both the underlying evidence and the resulting AI output.

The underlying input must be identifiable and preserved in its original form. Investigators should establish the source of the data, the device or system from which it was obtained, the date and method of acquisition, and the persons who handled it. Cryptographic hash values or equivalent integrity controls should be recorded before and after forensic processing. Any conversion, compression, enhancement,

segmentation, translation, or reconstruction must be documented.

The AI output must also be attributable to an identified system and analysis session. It must be possible to establish that the submitted result corresponds to the preserved input and the stated model version. Screenshots of an AI interface, unsupported statements that “the system found a match,” or manually copied results should not be sufficient. The output should be accompanied by machine-generated logs or a forensic report explaining how it was produced.

The legality of data acquisition must be examined independently from technical accuracy. AI analysis cannot cure an unlawful search, interception, seizure, or extraction. Where the input was obtained unlawfully or its authenticity cannot be established, Article 235(5) of the Criminal Procedure Code requires exclusion and removes its probative force. Where only a subsequent analytical step is defective, the original data may remain admissible, but the AI-generated interpretation may be excluded or accorded no evidentiary weight.

2) Scientific and Technical Reliability

The second layer determines whether the system is capable of performing the asserted forensic function reliably. The party introducing the evidence should bear the initial burden of presenting a sufficient technical foundation. This foundation should contain at least the system’s intended purpose, developer, model and software version, relevant validation studies, dataset characteristics, evaluation methodology, performance metrics, known limitations, and conditions of deployment.

Validation must correspond to the specific use. A system validated for verifying whether two controlled passport photographs depict the same person should not automatically be considered reliable for identifying a person from low-resolution night-time surveillance footage. A deepfake detector evaluated on English-language studio recordings should not automatically be treated as reliable for noisy Indonesian-language telephone audio. General product accuracy is

therefore not enough; the proponent must demonstrate fitness for the evidentiary purpose.

Performance reporting should identify false positives and false negatives rather than relying solely on aggregate accuracy. The appropriate metric depends on the evidentiary question. When a facial-recognition system searches a large database for a suspect, the risk of false association may be more significant than the risk of failing to identify the true person. When a deepfake detector is used to challenge an exculpatory recording, a false-positive classification could wrongly remove reliable defence evidence. Courts should therefore consider not only average performance but also the legal consequences of each type of error.

Dataset representativeness must be assessed. The evaluation data should reasonably reflect the relevant population, language, accent, recording quality, device, environmental conditions, and type of content. Where local validation has not been conducted, the proponent must explain why foreign or general validation results can be transferred to the Indonesian context. A lack of Indonesian validation does not invariably require exclusion, but it increases uncertainty and may require corroboration, restricted evidentiary use, or reduced weight.

The system should also be tested for robustness. The analysis must determine whether minor changes in compression, lighting, background noise, file format, cropping, or preprocessing materially affect the output. Where small and legally irrelevant modifications produce substantially different results, the system’s conclusion should not be treated as stable. Adversarial vulnerability is particularly relevant for synthetic-media detection because manipulated content may be deliberately designed to avoid detection.

Reliability further requires reproducibility or, at minimum, repeatability under documented conditions. An independent qualified examiner should be capable of running the preserved input through the same system version and obtaining a materially consistent output. Where the system is continuously updated, the version used in the case

must be preserved or made accessible. A later version should not be substituted because model updates may alter performance and outcomes.

Expert involvement remains necessary. The AI system does not replace an expert witness. The expert must explain the system's purpose, operation, validation, limitations, and interpretation and must remain responsible for the opinion presented to the court. Merely reading the system's output is not sufficient expertise. The expert should be able to explain why the system is appropriate for the particular data and why alternative explanations or sources of error were rejected.

3) Procedural Contestability

The third layer protects the accused's ability to challenge the evidence. Reliability that cannot be tested in the proceedings is legally incomplete. Due process and equality of arms require disclosure sufficient to permit the defence to examine the input, methodology, performance, and interpretation of the system.

The prosecution should disclose the original and processed input data; forensic acquisition records; hash values; chain-of-custody records; system name and version; configuration, thresholds, and material parameters; generated logs; validation and performance reports; known limitations; operator manuals; details of human intervention; and the expert's report. Relevant information concerning training or evaluation datasets should be disclosed at least in descriptive form, including their source, composition, representativeness, and limitations.

Trade-secret claims should be handled through protective measures rather than automatic nondisclosure. Courts may limit access to appointed experts, prohibit copying or secondary use, conduct closed technical hearings, or order redaction of commercially irrelevant information. Where withheld information is materially necessary to assess the reliability of prosecution evidence, courts should provide the defence with sufficient access to enable effective challenge while using proportionate protective

measures to safeguard legitimate commercial, privacy, or security interests. (Wexler, 2018) correctly observes that intellectual-property protection in criminal justice cannot be separated from liberty and fair-trial interests.

The defence must be afforded sufficient time and technical assistance. Providing thousands of pages of documentation immediately before trial would not constitute meaningful disclosure. When the prosecution relies materially on a complex AI system, the court should permit the defence to engage an independent expert. Where the accused lacks resources, publicly funded expert assistance should be available when necessary to prevent an inequality of arms.

A pretrial reliability hearing should be conducted when the AI output is disputed, technically novel, decisive, proprietary, or associated with a significant risk of error. At the hearing, the party introducing the evidence should establish its legal and technical foundation. The defence may question the system's validation, dataset, error rates, deployment conditions, and limitations. The judge should then determine the permitted use of the evidence before it is presented as reliable proof at trial.

3.4 Judicial Weight and Legal Consequences

The layered framework requires differentiated legal consequences. Failure of legal validity should ordinarily result in exclusion under Article 235(5). Where scientific reliability cannot be sufficiently established, the court should consider exclusion where the deficiency undermines the evidentiary foundation, while less substantial reliability concerns may instead justify limiting the evidentiary weight assigned to the AI-generated inference. Failure to provide meaningful disclosure should lead to postponement, compelled disclosure, restricted use, exclusion, or another remedy capable of restoring procedural fairness.

AI-derived inferences concerning identity, attribution, intention, or guilt should not constitute the sole or determinative basis of conviction (Allen, 2022). They should be

corroborated by independent evidence whose probative value does not depend on the same system, dataset, or analytical assumption. Corroboration is especially necessary where the model is proprietary, externally unvalidated, sensitive to input conditions, or associated with significant false-positive risks.

The court's judgment should expressly state the type and function of the AI evidence, the system and version used, the evidence on which the reliability finding was based, the principal limitations and error risks, the defence objections, the available corroborating evidence, and the weight assigned to the output. A judgment that merely describes an AI result as "sophisticated," "accurate," or "scientific" would be insufficient. Reasoned judicial assessment prevents the transfer of adjudicative responsibility from the judge to a technological system.

3.5 Institutional Implementation in Indonesia

The reconstructed framework should be implemented through procedural guidance rather than waiting for another comprehensive statutory amendment. The Supreme Court may issue a regulation or guideline concerning the examination of AI-based electronic evidence. The instrument should regulate disclosure, reliability hearings, expert qualifications, minimum forensic reports, judicial warnings, corroboration, and the consequences of noncompliance.

Law-enforcement institutions should establish a register of AI systems used in criminal investigation and forensic examination (Vashishtha, 2026). The register should record each system's developer, purpose, model version, validation status, approved uses, limitations, update history, security incidents, and responsible institutional unit. A system should not be deployed merely because it has been commercially purchased. Institutional validation and documented approval should precede its use in a criminal case.

Forensic laboratories using AI should implement quality-management procedures and independent auditing. Analysts should preserve

original data, document computational processing, participate in proficiency testing, and disclose uncertainty. Procurement agreements with private vendors should guarantee access to information necessary for judicial review. The State should not purchase a system under contractual terms that prevent courts or defendants from examining evidence used to restrict liberty.

The practical judicial sequence should consist of four stages. At the investigation stage, officers preserve the original data and computational chain of custody. At the prosecution stage, the prosecutor evaluates whether the system possesses sufficient legal and technical foundations and prepares the disclosure package. At the pretrial or preliminary stage, the court resolves disputes concerning legality, reliability, confidentiality, and defence access. At trial, the expert explains the result and is examined by both parties, while the judge assesses the output together with corroborating evidence.

This model does not reject technological innovation. It prevents technological efficiency from replacing evidentiary justification. AI can assist law enforcement in analysing large amounts of data, detecting manipulation, identifying investigative leads, and interpreting complex digital material. However, the closer an AI output comes to determining a disputed fact affecting guilt or liberty, the stronger the required standards of validation, transparency, independent examination, and judicial reasoning.

The proposed reconstruction consequently transforms the assessment of AI-based evidence from a formal question whether it may be classified as electronic evidence into a substantive and procedural inquiry concerning whether the evidence deserves to be trusted in the particular case. Through the cumulative examination of legal validity, scientific and technical reliability, and procedural contestability, Indonesian criminal procedure can use AI-based evidence while remaining consistent with material truth, the presumption of

innocence, equality of arms, and due process of law.

4. CONCLUSION

Indonesian criminal procedure recognizes AI-based evidence as electronic evidence, but existing rules do not adequately address its scientific reliability and procedural fairness. This study proposes a three-layer framework covering legal validity, scientific-technical reliability, and procedural contestability. AI evidence should be subject to enhanced scrutiny, independent corroboration, and reduced probative weight when reliability or contestability is inadequate. The study recommends Supreme Court guidelines on disclosure, reliability assessment, expert qualifications, forensic documentation, computational chain of custody, and corroboration. As an institutional safeguard, law-enforcement agencies should maintain documented records of AI systems used in criminal investigations, including their approved purposes, validation status, model versions, limitations, and update history, thereby supporting subsequent judicial verification. Through these safeguards, artificial intelligence may support the search for material truth without weakening the presumption of innocence, equality of arms, the right to defence, and due process of law.

AUTHOR CONTRIBUTIONS

Author 1: Writing & Original Draft.

Author 2: Methodology, Investigation, Formal, Analysis, Review & Editing.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Allen, D. N. (2022). Deepfake fight: AI-powered disinformation and perfidy under the Geneva Conventions. *Notre Dame J. on Emerging Tech.*, 3, 1.

Antoro, H., Budhijanto, D., & Somawijaya. (2026). Admissibility of artificial intelligence as electronic evidence: Comparative perspectives from Indonesia, the United States, and Japan. *Jurnal IUS Kajian Hukum dan Keadilan*, 14(1), 122–139.

<https://doi.org/10.29303/ius.v14i1.1880>

Biedermann, A., & Lau, T. (2024). Decisionalizing the problem of reliance on expert and machine evidence. *Law, Probability and Risk*, 23(1), Article mgae007.

<https://doi.org/10.1093/lpr/mgae007>

Chandra, B., Dunietz, J., Roberts, K., Lee, Y., Fontana, P., & Awad, G. (2024). *Reducing risks posed by synthetic content: An overview of technical approaches to digital content transparency* (NIST AI 100-4). National Institute of Standards and Technology.

<https://doi.org/10.6028/NIST.AI.100-4>

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations and Directives. *Official Journal of the European Union*, L, 2024/1689.

<https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>

Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face recognition vendor test part 3: Demographic effects* (NISTIR 8280). National Institute of Standards and Technology.

<https://doi.org/10.6028/NIST.IR.8280>

Fajar, M. N. D., & Achmad, Y. (2010). *Dualisme penelitian hukum normatif dan empiris*. Pustaka Pelajar.

Kattinig, M., Angerschmid, A., Reichel, T., & Kern, R. (2024). Assessing trustworthy AI: Technical and legal perspectives of fairness in AI. *Computer Law & Security*

- Review, 55, Article 106053.
<https://doi.org/10.1016/j.clsr.2024.106053>
- Lehr, D., & Ohm, P. (2017). Playing with the data: What legal scholars should learn about machine learning. *University of California, Davis Law Review*, 51(2), 653–717.
<https://lawreview.law.ucdavis.edu/archive/s/51/2/playing-data-what-legal-scholars-should-learn-about-machine-learning>
- Marzuki, P. M. (2017). *Penelitian hukum* (Edisi revisi). Kencana.
- National Institute of Justice. (2008). *Electronic crime scene investigation: A guide for first responders* (2nd ed.; NIJ Research Report, NCJ 219941). U.S. Department of Justice.
<https://nij.ojp.gov/library/publications/electronic-crime-scene-investigation-guide-first-responders-second-edition>
- Permadi, W. S., Isnandar, A., & Prasetyo, Y. (2026). The probative power of deepfake-based electronic evidence in fraud crimes. *Jurnal Hukum Prasada*, 13(1), 37–45.
<https://doi.org/10.22225/jhp.13.1.2026.37-45>
- Republic of Indonesia. (2019). Government Regulation Number 71 of 2019 concerning the operation of electronic systems and transactions.
<https://peraturan.bpk.go.id/Details/122030/pp-no-71-tahun-2019>
- Republic of Indonesia. (2024). *Law Number 1 of 2024 concerning the second amendment to Law Number 11 of 2008 concerning electronic information and transactions*.
<https://peraturan.bpk.go.id/Details/274494/uu-no-1-tahun-2024>
- Republic of Indonesia. (2025). *Law Number 20 of 2025 concerning the Criminal Procedure Code*.
<https://peraturan.bpk.go.id/Details/337302/uu-no-20-tahun-2025>
- Roth, A. (2017). Machine testimony. *The Yale Law Journal*, 126(7), 1972–2053.
<https://www.yalelawjournal.org/article/machine-testimony>
- Soekanto, S., & Mamudji, S. (2015). *Penelitian hukum normatif: Suatu tinjauan singkat*. RajaGrafindo Persada.
- State v. Loomis*, 881 N.W.2d 749 (Wis. 2016).
<https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690>
- State v. Pickett*, 246 A.3d 279 (N.J. Super. Ct. App. Div. 2021).
<https://www.njcourts.gov/court-opinion/state-of-new-jersey-vs-corey-pickett-17-07-0470-hudson-county-and-statewide-published>
- Sumardiana, B., Pujiyono, 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>
- Vashishtha, S. (2026). AI in cyber law enforcement and forensic evidence collection. *Scientific Journal of Artificial Intelligence and Blockchain Technologies*, 3(1), 13–23. <https://doi.org/10.63345>
- Wexler, R. (2018). Life, liberty, and trade secrets: Intellectual property in the criminal justice system. *Stanford Law Review*, 70(5), 1343–1429.
<https://www.stanfordlawreview.org/print/article/life-liberty-and-trade-secrets/>