Journal link: http://kmf-publishers.com/tatscr/

Theoretical and Applied Technological Science Review

Volume: 4 Issue: 4 Year: 2026
ISSN-2958-7824



KMF PUBLISHERS

www.kmf-publishers.com/tatscr/

Journal link: http://kmf-publishers.com/tatscr/

Received: 20 September 2026 | Revised: 25 October 2026 | Accepted: 28 October 2026

DOI: https://doi.org/10.64907/xkmf.v4i4.tatscr.2

Published: 30 October 2026

Research Article

Procedural Reliability Safeguards for AI Assisted Digital Evidence Analysis in Bangladesh

Mst. Dilara Islam Anny[1*]; Aungon Saha[1]; Md. Asraful Islam[1]; Kazi Abdul Mannan[2]

[1]Department of Law
[2]Department of Business Administration
Shanto-Mariam University of Creative Technology
Dhaka, Bangladesh

*Correspondence
Mst. Dilara Islam Anny
Email: adilaraislam@gmail.com

ABSTRACT

AI assisted digital evidence analysis requires safeguards that connect technical performance with legally examinable procedures. This study examines whether endorsement of independent validation and chain of custody is accompanied by endorsement of methodological documentation and human expert review in a Bangladesh based survey. The cross-sectional database contains 385 adult, AI aware respondents, including 253 law students, 67 law graduates and 32 advocates or legal professionals. Five digital evidence statements were analysed separately because their content combines perceived screening usefulness with distinct procedural requirements and does not establish a reflective reliability scale. Analyses comprise item distributions, six paired endorsement comparisons with Holm adjustment, joint response patterns, and two adjusted logistic models with robust covariance estimates. Independent validation was endorsed by 79.5% and chain of custody by 81.8%, compared with 59.0% for documentation and 63.4% for expert review. The four comparisons across these two pairs differed by 16.1-22.9 percentage points, with adjusted p values below .001. Only 29.4% endorsed all four safeguards. Validation endorsement was not clearly associated with documentation endorsement after adjustment (odds ratio 1.06, 95% interval 0.80-1.41) or review endorsement (0.94, 0.70-1.28). These findings identify uneven endorsement of complementary safeguards within this sample, rather than actual forensic reliability, national professional opinion or causal effects. Read alongside Bangladesh's digital record and expert evidence provisions, the results support education and reporting practices that connect validation and custody to inspectable methods and competent review. The study contributes an item-specific account of procedural reliability expectations while retaining explicit boundaries between survey attitudes, technical accuracy and legal admissibility.

Keywords: digital evidence; artificial intelligence; procedural safeguards; Bangladesh; expert evidence; protection motivation

Copyright: 2026 by the authors. Licensee KMF Publishers (www.kmf-publishers.com). This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1 Introduction

Digital evidence can pass through several stages before it becomes a proposition that a court is asked to accept. A device may be seized, an image created, messages extracted, files classified and an analyst's interpretation incorporated into a report. Artificial intelligence can enter at different points in this sequence, including sorting documents, identifying potentially relevant content or supporting interpretation. Its usefulness at one stage does not settle whether a later inference is dependable. A preserved file can be misclassified, while a technically promising classification can be difficult to scrutinise if its method and limitations are unavailable. Forensic guidance consequently treats collection, examination, analysis and reporting as connected but distinguishable activities.[1]

The governance problem is therefore more specific than whether people favour AI. Legal reliance requires reasons for accepting a particular output in a particular evidential setting. Computational reliabilism directs attention to technical and institutional grounds for justified reliance, extending beyond whether a system offers an explanation that appears understandable. A method must be considered in relation to its intended application and the circumstances of the individual case. An explanation can assist scrutiny, but intelligibility alone is not a demonstrated error rate, and an apparently confident answer is not an evidential foundation.[2]

Bangladesh provides an important setting for examining these distinctions. Section 65B of the Evidence Act 1872 addresses the admissibility of specified computer outputs subject to conditions concerning computer use, information input, operation and certification. Sections 45 and 45A address expert opinion and procedural requirements for physical or forensic expert evidence. These provisions create legally relevant questions about production, disclosure and expert assistance. They do not themselves establish the accuracy of every AI method that might be used on a digital record. Treating a statutory gateway as a general scientific endorsement would collapse different questions of proof.[3]

Research on AI adoption often asks whether perceived usefulness is associated with acceptance or trust. Digital evidence introduces an additional question: do respondents connect general safeguards with the practical conditions under which an output can be challenged? Independent validation, chain of custody, documentation and human review address different vulnerabilities. Validation concerns whether a method performs adequately for an

[1] Karen Kent and others, Guide to Integrating Forensic Techniques into Incident Response (NIST Special Publication 800-86, NIST 2006); James R Lyle and others, Digital Investigation Techniques: A NIST Scientific Foundation Review (NISTIR 8354, NIST 2022).

[2] Juan M Durán and others, 'From Understanding to Justifying: Computational Reliabilism for AI-Based Forensic Evidence Evaluation' (2024) 9 Forensic Science International: Synergy 100554.

[3] Evidence Act 1872 (Bangladesh), s 65B; Evidence Act 1872 (Bangladesh), s 45; Evidence Act 1872 (Bangladesh), s 45A.

intended use. Custody concerns continuity and handling. Documentation makes the relevant process inspectable. Review provides an opportunity for competent assessment of the output and its limitations. None is a complete substitute for the others.[4]

This distinction matters in legal education as well as professional practice. Students and graduates form part of the future institutional environment in which digital evidence will be requested, contested and evaluated. Their responses are relevant to educational planning and should be distinguished from the views of practising experts. The present database is predominantly student based: only 32 of its 385 respondents identify their current status as advocate or legal professional. A study that uses the full sample must therefore describe its mixed constituency and examine the smaller professional category separately, with suitable restraint about precision.

The questionnaire also requires close attention. Its digital evidence block contains one statement about organising or screening large volumes of evidence, followed by four statements about safeguards. Agreement that human review is necessary is not the same as confidence that current AI outputs are reliable. Likewise, demanding independent validation may express caution rather than trust. Averaging these statements into a single reliability score would obscure their substantive differences. Low internal consistency across the available five-item blocks further discourages an unqualified reflective measurement model, although internal consistency alone cannot decide what a construct means.[5]

The study consequently asks three connected questions. First, how frequently does this sample endorse screening usefulness and each procedural safeguard? Second, are validation and custody endorsed to the same degree as documentation and expert review within the same respondents? Third, is stronger validation or custody agreement associated with documentation and review endorsement after adjustment for selected observed characteristics? The third question is exploratory. Because all variables were recorded at one occasion, an adjusted relationship cannot establish which attitude came first or whether either caused the other.

The contribution lies in distinguishing endorsement levels from endorsement associations. A high aggregate proportion can coexist with weak alignment between individuals' responses to different safeguards. Joint response patterns make that possibility visible. Paired contrasts show whether the same respondents endorse statements differently, while adjusted models examine conditional covariation

[4] International Organization for Standardization and International Electrotechnical Commission, Information Technology—Security Techniques—Guidelines for Identification, Collection, Acquisition and Preservation of Digital Evidence (ISO/IEC 27037:2012, ISO 2012); International Organization for Standardization and International Electrotechnical Commission, Information Technology—Security Techniques—Guidelines for the Analysis and Interpretation of Digital Evidence (ISO/IEC 27042:2015, ISO 2015); Radina Stoykova, 'A New Right to Procedural Accuracy: A Governance Model for Digital Evidence in Criminal Proceedings' (2024) 55 Computer Law & Security Review 106040.

[5] Lee J Cronbach, 'Coefficient Alpha and the Internal Structure of Tests' (1951) 16 Psychometrika 297; Klaas Sijtsma, 'On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha' (2009) 74 Psychometrika 107.

rather than differences between marginal percentages. Together, these analyses provide a more specific account than a broad claim that AI literacy predicts evidential reliability. They also make an unsupported positive hypothesis interpretable without recasting uncertainty as proof of no relationship.[6]

The legal analysis then connects the observed pattern to evidential tasks without converting attitudes into legal rules. The survey cannot determine admissibility, validate an algorithm or establish a new statutory obligation. It can identify which elements of a defensible evidence process receive more or less explicit support in this sample. This narrower contribution is useful precisely because educational and institutional interventions need to know whether respondents connect a general demand for trustworthy AI with the documents, skills and procedures required for scrutiny. The following sections develop that argument through theory, transparent measurement, item-specific results and bounded implications.

2 Literature Review and Theoretical Foundation

The literature supplies three complementary perspectives on AI assisted evidence: technical grounds for reliability, procedural conditions for contesting evidence, and psychological accounts of protective judgement. This section integrates those perspectives around the actual questionnaire items. It distinguishes perceived usefulness from safeguard endorsement and develops exploratory expectations that can be examined without claiming that the survey measures tool accuracy or a complete latent theory.

2.1 Digital evidence and the scope of reliability

Digital evidence is not restricted to a final screenshot or printed message. Its probative significance can depend on how information was identified, acquired, preserved and interpreted. ISO/IEC 27037 addresses identification, collection, acquisition and preservation, while ISO/IEC 27042 addresses analysis and interpretation. Their separation is analytically useful: preservation of an object and justification of an interpretation are related tasks with different failure modes. These standards provide technical guidance; they are not treated here as automatically binding requirements of Bangladeshi evidence law.[7]

AI assistance likewise encompasses heterogeneous activities. Screening a large collection for potentially relevant files

[6] Quinn McNemar, 'Note on the Sampling Error of the Difference between Correlated Proportions or Percentages' (1947) 12 Psychometrika 153; Douglas G Altman and J Martin Bland, 'Absence of Evidence Is Not Evidence of Absence' (1995) 311 BMJ 485; Ronald L Wasserstein and Nicole A Lazar, 'The ASA Statement on p-Values: Context, Process, and Purpose' (2016) 70 American Statistician 129.

[7] ISO/IEC 27037:2012 (n 4); ISO/IEC 27042:2015 (n 4).

differs from attributing an account to a person or evaluating whether a recording is authentic. A statement that AI can help organise evidence expresses a perceived capability at the screening stage. It does not specify the model, training material, operating threshold, false positive rate or conditions of use. Accordingly, the present study uses “screening usefulness” for DER1 rather than treating it as a direct measurement of successful forensic performance.

Reliability also depends on the proposition for which a result is offered. A method may be useful for generating investigative leads while remaining insufficient for a strong claim of source identification. Durán and colleagues’ account of AI based forensic evaluation emphasises grounds for justified reliance across technical and broader scientific contexts. The implication for survey research is methodological: a respondent’s general confidence cannot replace evaluation of a specified method, application and case. A survey of expectations can illuminate governance priorities, but it cannot supply the validation evidence those expectations demand.[8]

2.2 Four complementary safeguards

Independent validation concerns evidence that a method is suitable for an intended task. Validation should be distinguished from the respondent merely agreeing that validation is desirable. A meaningful evaluation would identify reference material, relevant error patterns and the conditions under which performance deteriorates. The questionnaire does not ask respondents to assess any such evaluation. DER2 therefore captures a requirement for validation, not the existence or quality of a validation study. NIST’s scientific foundation review offers a broader basis for examining digital investigation techniques without implying that all methods share the same evidential maturity.[9]

Chain of custody concerns the continuity and handling of evidential material. It can help establish which object was acquired, who handled it and whether a working copy corresponds to a preserved source. These questions remain important when AI tools are introduced because model outputs may become additional artefacts requiring traceability. Nevertheless, a complete custody record cannot establish that the model’s inference was correct. The present study therefore treats DER3 as a distinct continuity requirement, not an alternative measure of algorithmic accuracy.[10]

Documentation supports scrutiny by making relevant aspects of a process available to

[8] Durán and others (n 2).
[9] Lyle and others (n 1).
[10] Kent and others (n 1); ISO/IEC 27037:2012 (n 4).

another person. In an AI assisted workflow, useful documentation may include the task, system version, inputs, preprocessing, operating settings, output and known limitations. Model cards and internal auditing frameworks illustrate approaches to recording intended uses and accountability throughout development and deployment. These approaches inform the practical discussion, but the survey does not measure whether respondents have encountered or prepared such records. DER4 concerns agreement that methods should be documented sufficiently for scrutiny.[11]

Human expert review introduces judgement into the assessment of an AI supported interpretation. Review is meaningful only when the reviewer has appropriate expertise, access to the necessary information and an opportunity to challenge the output. Green's analysis of human oversight policies warns against assuming that the presence of a person automatically resolves automated decision problems. Thus DER5 should not be interpreted as a preference for an infallible human decision maker. It asks whether expert review is necessary for high-stakes interpretation, leaving the competence and effectiveness of that review unmeasured.[12]

2.3 Protection motivation as an interpretive foundation

Protection Motivation Theory explains protective responses through threat and coping appraisal. The revised account developed by Maddux and Rogers includes self-efficacy alongside other appraisals relevant to protective action, and subsequent meta-analysis examines the wider empirical record of the theory. In a digital evidence setting, concern about error may make safeguards attractive, while beliefs about the feasibility and effectiveness of protective procedures may affect whether those safeguards are supported or implemented.[13]

The available questionnaire does not operationalise the entire theory. It contains no matched, validated set of severity, vulnerability, response cost, response efficacy and self-efficacy measures for performing a specified forensic task. It also records agreement rather than observed protective action. Protection motivation is therefore used to interpret why related safeguards might be endorsed together, rather than as a complete structural model. A full theoretical test requires direct measurement of the relevant threat and coping appraisals.

The theory nevertheless sharpens a useful distinction. General concern may support visible protective requirements such as validation or preservation, while an operational requirement such as

[11] Margaret Mitchell and others, 'Model Cards for Model Reporting' in Proceedings of the Conference on Fairness, Accountability, and Transparency (ACM 2019) 220; Inioluwa Deborah Raji and others, 'Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing' in Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (ACM 2020) 33.

[12] Ben Green, 'The Flaws of Policies Requiring Human Oversight of Government Algorithms' (2022) 45 Computer Law & Security Review 105681.

[13] James E Maddux and Ronald W Rogers, 'Protection Motivation and Self-Efficacy: A Revised Theory of Fear Appeals and Attitude Change' (1983) 19 Journal of Experimental Social Psychology 469; Donna L Floyd, Steven Prentice-Dunn and Ronald W Rogers, 'A Meta-Analysis of Research on Protection Motivation Theory' (2000) 30 Journal of Applied Social Psychology 407.

documentation may require a more concrete understanding of what a practitioner must do. The survey cannot establish that this mechanism explains its results, but it can assess whether endorsement aligns across the relevant statements. If validation agreement were positively associated with documentation and review after adjustment, that would be compatible with a more integrated protective orientation. If the association were uncertain, the theory would remain open to further investigation rather than being disproved.

2.4 Trustworthiness and procedural scrutiny

Organisational trust theory distinguishes characteristics of a trustee from willingness to accept vulnerability. This distinction helps prevent a common interpretive error in AI surveys: requirements for trustworthy conduct are not themselves expressions of willingness to rely on a system. Jacovi and colleagues similarly examine the conceptual conditions under which trust in AI should be understood. The present items address desired safeguards, not a direct trust scale or a behavioural decision to accept an evidential conclusion.[14]

Procedural scrutiny supplies the institutional counterpart to this measurement boundary. Stoykova's account of procedural accuracy develops a governance perspective on digital evidence in criminal proceedings. Its comparative legal setting cannot be imported as a binding Bangladeshi right, but its attention to accuracy and procedural protection helps explain why an evidential output must remain open to examination. Documentation and competent review connect technical claims with a process in which parties can question how an inference was produced.[15]

Bangladesh's statutory framework reinforces the need to distinguish these issues carefully. Section 65B specifies conditions under which a computer output may be admitted without producing the original, while section 45 addresses the relevance of expert opinion. Section 45A provides for disclosure of a physical or forensic expert report to the parties before testimony, subject to the court's power to grant leave, and identifies the expert's duty to assist the court. These provisions address different aspects of the evidence process. The present analysis does not infer that statutory compliance alone validates an AI model or makes its conclusion conclusive.[16]

[14] Roger C Mayer, James H Davis and F David Schoorman, 'An Integrative Model of Organizational Trust' (1995) 20 Academy of Management Review 709; Alon Jacovi and others, 'Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI' in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (ACM 2021) 624.

[15] Stoykova (n 4).

[16] Evidence Act 1872 (Bangladesh), s 65B; Evidence Act 1872 (Bangladesh), s 45; Evidence Act 1872 (Bangladesh), s 45A.

2.5 Exploratory expectations and analytical framework

The first expectation concerns alignment between validation and operational safeguards. Independent validation can supply grounds for evaluating performance, but those grounds become difficult to scrutinise if the method is undocumented. Expert review similarly requires information about how an output was generated and what limitations apply. Reliability and accountability scholarship therefore provides a reason to expect positive covariation between validation agreement and endorsement of both documentation and review. This is a theory-informed expectation, not evidence that respondents consciously follow a single reasoning sequence.[17]

H1: Stronger agreement with independent validation is positively associated with endorsement of methodological documentation and human expert review, conditional on the specified adjustments.

The second expectation concerns custody. Continuity practices can encourage attention to evidential traceability, which may also support documentation and review. However, continuity and interpretive scrutiny remain separate requirements. A respondent might be strongly familiar with preservation language without appreciating how model versioning, settings or error assessment affect an AI supported inference. The expected association is consequently plausible but not inevitable. The statistical model tests whether the measured relationship is sufficiently clear in these records, rather than treating normative complementarity as proof of psychological coherence.[18]

H2: Stronger agreement with chain of custody is positively associated with endorsement of methodological documentation and human expert review, conditional on the specified adjustments.

The hypotheses were developed for this reanalysis of an existing questionnaire and are exploratory rather than preregistered predictions. Both are examined across two outcomes. Perceived screening usefulness and self-reported verification are included as adjustments because a generally favourable orientation toward AI and reported checking behaviour might accompany safeguard responses. Training, current professional status, use frequency and age provide limited contextual adjustments. None is presented as an exhaustive account of background differences, and the coefficients should not be interpreted as causal effects.

Figure 1. Exploratory association framework. H1 links validation and H2 links custody to documentation and expert review. Arrows identify model terms, not causal paths. Actual accuracy and willingness to rely are not measured.

[17] Durán and others (n 2); Raji and others (n 11).
[18] Kent and others (n 1); ISO/IEC 27042:2015 (n 4).

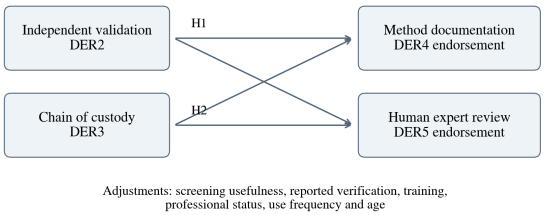


Figure 1. Exploratory association framework

Figure 1 separates the measured associations from the unmeasured concepts that motivate the study. There is no mediator or moderator because the data do not support a defensible temporal mechanism. The framework also avoids placing actual forensic reliability at the end of a statistical path: no experimental accuracy outcome was collected. Its purpose is to ask whether different components of a procedurally defensible workflow are endorsed together, while retaining an explicit distinction between a desirable requirement and its successful implementation.

3 Methodology

The study uses a cross-sectional quantitative design with item-specific descriptive and exploratory association analyses. The unit of analysis is one eligible survey record. The analytical strategy follows the wording of the instrument rather than assuming that each labelled questionnaire block forms a validated latent construct. This section documents the sample, measurement boundaries, screening decisions and uncertainty estimates so that the reported results can be reproduced and assessed.

3.1 Research design

The analysis examines the Group 6 Law-SDG survey database for Summer 2026. It combines survey analysis with a doctrinal reading of the relevant digital record and expert evidence provisions. The legal material provides context for interpreting procedural requirements; it is not another quantitative data source and does not turn the questionnaire into a test of statutory compliance. All associations are contemporaneous. Terms such as effect, impact and determinant are therefore avoided when describing regression coefficients. The study neither evaluates an AI product nor asks participants to decide an actual evidentiary dispute.

3.2 Study population and context

The records concern adults who reported AI awareness in a Bangladesh based legal research setting. Eligibility fields record affirmative consent, adult status and awareness for all 385 respondents. Current status includes law students, law graduates, advocates or legal professionals, and another relevant respondent category. The latter category is not further differentiated. The full sample is described as mixed rather than professional. Education level is a separate field; its “Professional” category is not substituted for current advocate or legal professional status. Geographic and experience fields are treated as respondent characteristics rather than evidence of national coverage.

3.3 Sampling and sample size

The final available sample comprises 385 complete numerical records. Recruitment dates, a sampling frame, invitation counts and recruitment probabilities are not recorded in the analytical dataset. Accordingly, the study does not assert probability sampling, calculate a response rate or claim that the sample represents all Bangladeshi legal professionals. A count of 385 is numerically close to the familiar large-population calculation for a proportion at 95% confidence, maximum variability and five-percentage-point precision. That arithmetic does not establish that those design assumptions were satisfied and is not used as retrospective evidence of representativeness.[19]

All available eligible records were retained in the primary analysis. The 32 respondents currently identifying as advocates or legal professionals form a descriptive subgroup; the dataset does not distinguish judges, prosecutors, defence lawyers or forensic specialists. Small subgroup size limits precision and discourages an adjusted professional-only model. The subgroup results are intended to make sample composition visible, not to substitute a small professional estimate for a national assessment. Statistical intervals elsewhere are conditional on independent-record assumptions and cannot correct unknown selection or recruitment clustering.

3.4 Questionnaire content and measurement

The workbook contains demographic and AI-use fields, four five-item agreement blocks and additional tool and understanding fields. Agreement items range from one, strongly disagree, to five, strongly agree, with three neutral. DER1 concerns screening usefulness. DER2-DER5 concern independent validation, custody, documentation and expert review. AIL2 records self-reported verification of important AI generated information. These statements do not directly measure objectively demonstrated literacy, current institutional compliance or actual forensic accuracy. Table 1 maps their analytical roles; Appendix A reproduces their exact wording.

Table 1. Item content and analytical roles

Code	Abbreviated content	Analytical role
DER1	AI can organise or screen evidence	Perceived capability; adjustment
DER2	Independent validation before reliance	Safeguard; focal predictor
DER3	Chain of custody remains essential	Safeguard; focal predictor
DER4	Methods documented for scrutiny	Safeguard; outcome
DER5	Expert review for high-stakes interpretation	Safeguard; outcome
AIL2	Verify important AI information	Reported behaviour; adjustment

Source: Group 6 Codebook sheet. Exact focal wording appears in Appendix A. These are separate statements, not indicators of a validated reflective reliability scale.

Content interpretation precedes aggregation. The DER items combine a capability proposition with normative requirements,

[19] William G Cochran, Sampling Techniques (3rd edn, John Wiley & Sons 1977).

while the five-item blocks show weak internal consistency in these data. Cronbach's alpha is reported diagnostically, not as a certificate of validity or as a reason to delete inconvenient items. No composite reliability, average variance extracted or discriminant-validity statistic is claimed for a reflective model that the measurement content does not justify. The primary outcome definition is endorsement, meaning a response of four or five. Nonendorsement includes neutrality as well as disagreement.[20]

3.5 Instrument refinement and pilot evidence

No separate pilot dataset, cognitive interview record or item revision log accompanies the analytical materials. A pilot sample size, expert review panel or completed translation-validation procedure is therefore not asserted. The available items permit an exploratory study of specific opinions, but further instrument development would be needed for a latent scale of procedural reliability orientation. The current analysis preserves the observed wording and avoids post hoc item removal to improve alpha. This approach retains an interpretable connection between each reported result and the question participants answered.

3.6 Data collection record and screening

The researchers' consent and participant-protection account is retained in Section 8. The numerical audit checks eligibility, missingness, permissible values, identifier duplication and response patterns. No agreement response is missing or outside the one-to-five range. There are no duplicate identifiers, duplicate full 20-item response vectors or records giving the same answer to all 20 agreement items. These checks do not establish identity, attentive participation or response authenticity individually. No separate attention-check variable was available. Sixteen tool-name entries are missing; they are not required for the models.

Experience and frequency fields contain discordant "Never" responses: 20 records report never using AI by experience but some use by frequency, and 14 show the reverse pattern. Primary analysis retains them because item responses remain available and the discrepancy may reflect differing interpretations of experience and frequency. A sensitivity analysis excludes either form of discordance, leaving 351 records. The contextual-text field is excluded throughout. Its contents do not contribute to recruitment, interpretation, illustrative quotations or the computation of results.

[20] Cronbach (n 5); Sijtsma (n 5).

3.7 Analytical strategy

Item distributions include means, standard deviations, category counts and endorsement proportions with Wilson 95% intervals. All six paired comparisons among DER2-DER5 use exact McNemar tests and a single six-test Holm correction. Percentage-point differences receive pointwise percentile intervals from 10,000 whole-record bootstrap resamples, using seed 20261025. Resampling complete records preserves the dependence between answers from the same respondent. These intervals quantify model-based uncertainty within the records, not uncertainty from a documented probability design.[21]

Two logistic models examine endorsement of documentation and review. Each includes validation, custody, screening usefulness, self-reported verification, any formal training, advocate or legal professional status, frequency and age 27 or above. Four agreement predictors are centred at three; frequency is coded zero to four from never to daily and centred at monthly. HC1 sandwich covariance uses the n/(n−k) correction, with k including the intercept. H1 and H2 each have a separate two-outcome Holm family. Intervals remain pointwise. Sensitivity analyses restrict to consistent-use records or current law students and redefine the outcome as strong agreement. Estimation was independently cross-checked using a second logistic implementation. No mediation, causal identification or out-of-sample prediction is claimed.[22]

3.8 Ethical considerations

Formal ethical approval was not obtained in the institutional circumstances described in the authors' declaration. That circumstance is reported as context rather than presented as an ethical exemption or external finding that review was unnecessary. The researchers retained responsibility for participant protection. Participation required adult eligibility and affirmative informed consent; respondents were informed of the purpose, anonymity, voluntary participation and opportunity to discontinue before questionnaire completion. All analysed records meet the recorded consent and age conditions.

The analytical workflow uses respondent codes and aggregate reporting. It excludes contextual comments, does not seek to identify participants and reports the small professional group only through broad category summaries. Data sharing remains subject to confidentiality and institutional requirements, as stated in Section 8. Withdrawal during completion should be distinguished from later removal of a truly anonymous submission, for which identification may be impossible. The manuscript does not promise a post-submission identification mechanism that is absent from the data. These safeguards explain the conduct of the research without implying that strong protection practices retrospectively create formal approval.

[21] Edwin B Wilson, 'Probable Inference, the Law of Succession, and Statistical Inference' (1927) 22 Journal of the American Statistical Association 209; McNemar (n 6); Sture Holm, 'A Simple Sequentially Rejective Multiple Test Procedure' (1979) 6 Scandinavian Journal of Statistics 65; Bradley Efron and Robert J Tibshirani, An Introduction to the Bootstrap (Chapman & Hall 1993).

[22] Leonard A Stefanski and Dennis D Boos, 'The Calculus of M-Estimation' (2002) 56 American Statistician 29; Holm (n 21); Wasserstein and Lazar (n 6).

4 Results

The results move from sample composition to item-specific responses, paired contrasts and exploratory adjusted associations. This sequence distinguishes what respondents endorsed from whether their endorsements covaried. Unless a subgroup is explicitly identified, the denominator is 385. Confidence intervals and p values refer to the stated analytical models; they do not confer population representativeness on the sample.

4.1 Respondent profile

Law students account for 253 records (65.7%), law graduates for 67 (17.4%), advocates or legal professionals for 32 (8.3%), and other relevant respondents for 33 (8.6%). Most respondents are below 27 years of age: 304 records (79.0%). Men comprise 193 records and women 179; nine respondents prefer not to disclose gender and four select another category. These characteristics establish the predominantly early-career composition of the sample before any substantive inference is considered.

Table 2. Respondent profile in the complete database

Characteristic	Categories and counts
Age	21-23: 151; 24-26: 85
	18-20: 68; 31+: 48
	27-30: 33
Gender	Male: 193; Female: 179
	Prefer not to say: 9; Other: 4
Current status	Law student: 253; Law graduate: 67
	Other relevant respondent: 33; Advocate or legal professional: 32
Education	Undergraduate: 221; Graduate: 80
	Postgraduate: 42; Professional: 42
Location	Dhaka: 150; District town: 87
	Other city corporation: 79; Upazila-rural area: 59
	Other: 10
Experience	1-2 years: 130; 6-12 months: 111
	Less than 6 months: 65; More than 2 years: 59
	Never: 20
Frequency	Weekly: 159; Daily or almost daily: 120
	Monthly: 63; Less than monthly: 29
	Never: 14
Training	No: 224; Yes, short orientation: 113
	Yes, course or certified training: 48

N = 385. Counts are unweighted. Current professional status and professional education are distinct fields. Categories reproduce the database labels.

Dhaka contributes 150 records (39.0%), while the remaining categories include district towns, other cities, rural or upazila locations and other settings. The database records no formal AI training for 224 respondents (58.2%); 113 report a short orientation and 48 a course or certification. Frequency is weekly for 159 and daily or almost daily for 120. These are self-reported use categories, not evidence of experience with forensic software specifically. The 42 respondents selecting professional education

must not be confused with the 32 selecting current advocate or legal professional status.

4.2 Item distributions and measurement diagnosis

Independent validation and custody attract the highest endorsement, at 306/385 (79.5%) and 315/385 (81.8%). Documentation is endorsed by 227 (59.0%), review by 244 (63.4%), and screening usefulness by 223 (57.9%). Table 3 reports the underlying five-category counts so that differences are not hidden by the binary threshold. Validation and custody also have higher means, but the agreement scale should not be treated as a directly calibrated quantity of actual reliability.

Table 3. Digital evidence item distributions

Item abbreviated	Mean (SD)	Counts 1 / 2 / 3 / 4 / 5	Endorsed n (%)	95% interval %
DER1: Screening usefulness	3.67 (0.84)	1 / 27 / 134 / 160 / 63	223 (57.9%)	52.9-62.8
DER2: Validation	4.11 (0.75)	0 / 5 / 74 / 180 / 126	306 (79.5%)	75.2-83.2
DER3: Custody	4.10 (0.73)	0 / 8 / 62 / 200 / 115	315 (81.8%)	77.7-85.4
DER4: Documentation	3.68 (0.82)	0 / 25 / 133 / 166 / 61	227 (59.0%)	54.0-63.8
DER5: Expert review	3.76 (0.78)	0 / 17 / 124 / 180 / 64	244 (63.4%)	58.5-68.0

N = 385. Endorsed means agree or strongly agree (four or five). Three is neutral. Intervals are pointwise Wilson intervals and do not establish population representativeness.

Neutral responses are substantively important. Documentation has 133 neutral responses and review 124; validation has 74 and custody 62. Thus the lower endorsement of documentation and review is not principally evidence of explicit opposition. Only 25 respondents disagree with documentation and 17 disagree with review; none strongly disagrees with either statement. Describing all responses below four as rejection would overstate the pattern. The result concerns weaker affirmative endorsement, with substantial uncertainty or neutrality about some requirements.

The five-item alpha estimates are 0.311 for AI literacy statements, 0.246 for cyberattack-prevention statements, 0.266 for the digital evidence block and 0.204 for deepfake-related statements. These values do not support treating the blocks as well-established reflective scales in this sample. More importantly, the digital evidence items already have different substantive referents. The main analysis therefore does not estimate paths between block averages or present them as validated psychological constructs. The low values remain a measurement finding rather than being concealed by selective deletion.

4.3 Paired differences in safeguard endorsement

The same records supply all four safeguard responses, so paired comparisons are more informative than independent-proportion tests. Table 4 displays all six comparisons, including the comparisons that do not show clear differences. Validation exceeds documentation by 20.5 percentage points and review by 16.1 points. Custody exceeds documentation by 22.9 points and review by 18.4 points. Each of these four comparisons has a Holm-adjusted p value below .001.

Table 4. All paired safeguard endorsement comparisons

First minus second	First only / second only	Gap pp	95% interval pp	Holm p
Validation – custody	52 / 61	-2.3	-7.8-2.9	0.452
Validation – documentation	121 / 42	20.5	14.5-26.8	< .001
Validation – review	108 / 46	16.1	10.1-22.3	< .001
Custody – documentation	126 / 38	22.9	16.6-28.8	< .001
Custody – review	111 / 40	18.4	12.5-24.4	< .001
Documentation – review	76 / 93	-4.4	-10.9-2.1	0.437

N = 385. pp = percentage points. Intervals use 10,000 whole-record bootstrap resamples. Exact McNemar tests use discordant responses; Holm adjustment covers all six tests. Intervals are pointwise.

The validation-minus-custody contrast is –2.3 points, with a 95% bootstrap interval from –7.8 to 2.9 and adjusted p = .452. The documentation-minus-review contrast is –4.4 points, with an interval from –10.9 to 2.1 and adjusted p = .437. The data therefore distinguish two pairs of endorsement levels more clearly than they establish a strict four-item ranking. Differences in wording and perceived task familiarity remain possible explanations; the questionnaire does not directly compare the relative importance of safeguards under a common scenario.

4.4 Joint patterns of safeguard support

Among all respondents, 198 endorse both validation and review, 108 endorse validation without review, 46 endorse review without validation and 33 endorse neither. Table 5 makes these joint responses visible. Among the 306 validation endorsers, 108 (35.3%) do not also endorse review. This conditional proportion includes neutral responses and must not be described as the proportion opposed to experts.

Table 5. Joint endorsement of validation and expert review

Validation endorsed	Review endorsed	n (%)
Yes	Yes	198 (51.4%)
Yes	No	108 (28.1%)
No	Yes	46 (11.9%)
No	No	33 (8.6%)

N = 385. No includes neutrality and disagreement. All-four safeguard endorsement is 113/385 (29.4%). Among validation endorsers, 108/306 (35.3%) do not endorse review.

All four safeguards are endorsed by 113 respondents (29.4%). Another 134 endorse three, 101 endorse two, 36 endorse one and one endorses none. These counts describe the conjunction of four distinct statements rather than a validated additive scale. They show that widespread support for a prominent requirement can coexist with substantially less common endorsement of the complete set. No inference is made that each nonendorser would refuse a safeguard in a real case; behavioural choices were not observed.

4.5 Adjusted associations

The two logistic models retain all 385 records and contain eight non-intercept predictors. Table 6 reports validation and custody coefficients; Appendix C provides the remaining coefficients. For documentation, the validation odds ratio is 1.06 (95% interval 0.80-1.41). For review it is 0.94 (0.70-1.28). Their separate two-outcome Holm family yields adjusted p = 1.000 for each. H1 therefore receives no clear support in these data.

Table 6. Adjusted validation and custody associations

Outcome	n	Validation OR (95% CI)	Holm p	Custody OR (95% CI)	Holm p
Documentation	227	1.06 (0.80-1.41)	1.000	1.14 (0.85-1.54)	0.595
Expert review	244	0.94 (0.70-1.28)	1.000	1.17 (0.87-1.58)	0.595

N = 385 in each model; n is the endorsed outcome count. OR = odds ratio per score point; CI = pointwise HC1 interval. Each focal predictor has a separate two-outcome Holm family. Models include all eight non-intercept predictors in Appendix C.

Custody agreement has odds ratios of 1.14 (0.85-1.54) for documentation and 1.17 (0.87-1.58) for review, with Holm-adjusted p = .595 for both. H2 likewise receives no clear support. These estimates allow both relatively small and appreciable positive associations within their intervals; they do not establish practical equivalence. The maximum variance inflation factor is 1.07, so substantial linear collinearity among the specified predictors is not apparent. Independent estimation reproduces coefficients within 0.000001, supporting computational consistency without resolving design or measurement limitations.

4.6 Professional subgroup and sensitivity analyses

The advocate or legal professional subgroup shows endorsement of validation by 26/32 (81.3%), custody by 30/32 (93.8%), documentation by 19/32 (59.4%) and review by 18/32 (56.3%). Screening usefulness is endorsed by 13/32 (40.6%). Table 7 reports Wilson intervals because these small counts carry considerable uncertainty. The pattern is descriptively compatible with higher support for validation and custody, but a precise comparison with all other respondents is beyond the subgroup's intended analytical role.

Table 7. Advocate or legal professional subgroup

Statement	Endorsed n (%)	95% interval %
Screening usefulness	13 (40.6%)	25.5-57.7
Validation	26 (81.3%)	64.7-91.1
Custody	30 (93.8%)	79.9-98.3
Documentation	19 (59.4%)	42.3-74.5
Expert review	18 (56.3%)	39.3-71.8

n = 32. Intervals are pointwise Wilson intervals. These descriptive estimates are not nationally representative and do not distinguish legal occupations or forensic specialisms.

The professional estimates should not be read as separate findings for particular legal occupations. No subgroup category identifies the number of practising forensic experts or respondents who have personally used AI in evidence analysis. The subgroup's lower screening usefulness endorsement is consequently not evidence that professional experience causes scepticism. It is a descriptive result that future occupationally stratified research could examine with task-specific measures.

Three sensitivity specifications examine the main validation coefficients. Restricting to 351 consistent-use records gives odds ratios of 1.07 for documentation and 1.00 for review. Restricting to 253 current law students gives 1.21 and 0.88. Using strong agreement as the outcome gives 1.33 and 1.10. Every corresponding interval includes one. Table 8 presents the interval estimates rather than using threshold changes to search for a favourable result.

Table 8. Sensitivity of adjusted validation associations

Specification	n	Documentation OR (95% CI)	Review OR (95% CI)
Neither use field says never	351	1.07 (0.79-1.44)	1.00 (0.72-1.38)
Current law students	253	1.21 (0.86-1.71)	0.88 (0.60-1.28)
Strong agreement outcome	385	1.33 (0.86-2.06)	1.10 (0.77-1.59)

All estimates concern DER2. The student model omits invariant professional status. Strong agreement models contain 61 documentation and 64 review events. Intervals use HC1 covariance; specifications estimate different outcomes or constituencies.

These analyses do not reveal a clearly positive validation association under the alternative definitions. They also do not demonstrate that all specifications estimate the same quantity: the strict outcome concerns strong agreement, while subgroup restrictions change the analysed constituency. Taken together, the results support a descriptive conclusion of uneven safeguard endorsement and an association conclusion of uncertainty. They do not support a causal claim that validation attitudes increase or decrease documentation or review expectations.

Figure 2. Paired safeguard endorsement contrasts and adjusted validation associations. Intervals are pointwise 95% bootstrap intervals in panel A and HC1 intervals in panel B. N = 385.

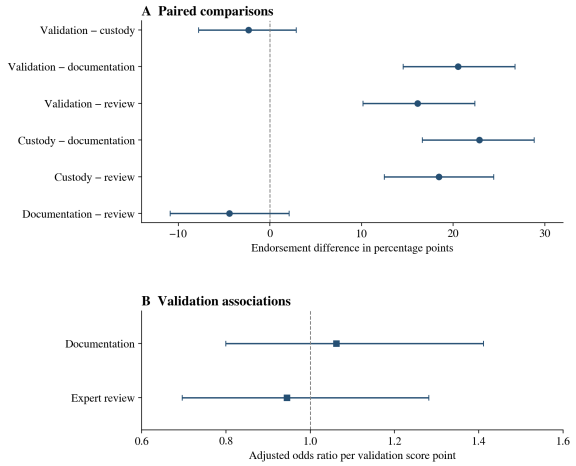


Figure 2. Paired safeguard endorsement contrasts and adjusted validation associations

Figure 2 presents the paired contrasts alongside the adjusted validation estimates. The visual separation is deliberate: a difference between endorsement proportions and an odds ratio describing individual-level association answer different questions. The former is clear for the four comparisons across the two pairs; the latter remains imprecise around an odds ratio of one. Neither should be used as a substitute for the other when summarising the findings.

5 Discussion

The principal finding is an uneven pattern of explicit support for complementary procedural safeguards. Validation and custody receive substantially more endorsement than documentation and expert review, yet stronger agreement with validation is not clearly associated with endorsement of either operational safeguard after adjustment. This combination suggests that aggregate support for trustworthy AI should not be treated as evidence that respondents connect all the procedures required for scrutiny. The following interpretation remains bounded by what the questionnaire measured and by the predominantly student composition of the sample.

5.1 Validation and custody as visible requirements

High endorsement of validation and custody is consistent with a view that AI supported evidence needs an externally defensible foundation and a traceable history. These are reasonable requirements, but they protect different elements of an evidential claim. Validation concerns performance for an intended application, while custody concerns continuity of evidential material. NIST and ISO guidance make that distinction visible through separate attention to acquisition, preservation, analysis and reporting. The survey's strongest endorsements therefore concern important components of reliability governance, although they provide no evidence that the components are actually implemented.[23]

The paired results do not imply that custody is intrinsically more important than review or that validation should receive greater legal priority than documentation. Respondents were not asked to allocate limited resources or rank safeguards for an identical case. Different statement wording can elicit different levels of familiarity or agreement. What can be concluded is narrower: in these records, affirmative responses are consistently more common for the validation and custody statements than for documentation and review. This pattern identifies an educational question about how familiar safeguards connect to the practical production of inspectable evidence.

[23] Kent and others (n 1); Lyle and others (n 1); ISO/IEC 27037:2012 (n 4); ISO/IEC 27042:2015 (n 4).

5.2 Documentation as the link to scrutiny

Documentation has the lowest endorsement of the four safeguards, with more than a third of respondents neutral. That matters because a general demand for validation becomes difficult to examine without information about what was validated, under which conditions and for what intended use. Durán and colleagues' account of justified forensic reliance supports attention to such grounds rather than confidence in an unexplained output. Stoykova's procedural perspective similarly draws attention to the conditions under which digital evidence can be evaluated within a fair process. The present result is compatible with a gap in explicit articulation of those conditions; it does not prove that respondents lack the underlying knowledge.[24]

Practical documentation need not disclose every source-code component to every participant. The relevant question is which information is necessary for meaningful scrutiny of the particular claim. A model identifier, input history, processing record and account of limitations may each be important, depending on the task. Model cards and auditing frameworks offer examples of structured disclosure and organisational accountability, but they do not themselves answer every evidential question. The study supports examining documentation requirements at the level of the claim and workflow rather than treating "AI transparency" as a single, self-executing solution.[25]

5.3 Expert review without automatic deference

Review is endorsed more frequently than documentation but less frequently than validation or custody. The joint pattern is particularly informative: 108 validation endorsers do not affirm the review requirement. This does not show hostility toward experts because neutrality is included. It shows that validation endorsement cannot be assumed to include a demand for expert interpretation in every respondent's answers. The distinction should shape how training presents the relationship between method evaluation and case-specific judgement.

Human oversight must also be specified carefully. A reviewer who sees only a summary may be unable to detect an inappropriate input, an unsupported attribution or a performance limitation outside the validation setting. Green's critique of oversight policies cautions that an institutional requirement to place a person in a workflow does not establish meaningful control. The appropriate response is therefore not a general instruction to trust humans instead of machines. It is to define the expertise, information, authority and reasoning expected of the reviewer, and to make those conditions available for examination.[26]

[24] Durán and others (n 2); Stoykova (n 4).
[25] Mitchell and others (n 11); Raji and others (n 11).
[26] Green (n 12); Lyle and others (n 1).

5.4 Interpreting the unsupported associations

The adjusted models do not support the hypothesised positive relationships clearly. A respondent can demand validation while remaining neutral about documentation or review, and the aggregate difference does not require a strong person-level association. The model intervals also allow a range of effects, including moderate positive associations. Reporting only nonsignificant p values would hide that uncertainty. The ASA statement and the distinction between absence of evidence and evidence of absence support interpreting the estimates with their intervals and design constraints.[27]

Several explanations remain possible. Restricted response variation at the upper end of validation and custody may limit association estimates. The statements may elicit separate normative judgements rather than a common protective orientation. Respondents may differ in task interpretation, and the adjustments do not capture forensic experience or institutional resources. Protection Motivation Theory provides a possible account involving threat and coping appraisal, but those mechanisms were not adequately measured here. The current results therefore constrain the proposed integration without demonstrating that protection motivation is absent or theoretically irrelevant.[28]

5.5 Bangladesh law and the meaning of legal reliance

Section 65B's computer-output conditions make record production and certification legally salient, but they should not be conflated with the validation of an AI analytical method. A certificate identifying a record and describing its production can assist the statutory inquiry while leaving disputed questions about classification, inference or model error unresolved. The distinction between the admissibility of a record and the strength of an inference drawn from it is essential to interpreting the present results. High endorsement of custody does not demonstrate understanding of every condition in section 65B, because the survey did not test that knowledge.[29]

Sections 45 and 45A place expert opinion and forensic reporting within a judicial process. The disclosure requirement and duty to assist the court provide reasons to connect expert review with an intelligible account of method and limitations. They do not authorise a survey-based conclusion that every AI output requires precisely the same expert intervention. Nor does this study determine how a court would resolve a particular dispute. Its practical contribution is to identify a pattern of attitudes relevant to preparing people to ask the right evidential questions.[30]

[27] Wasserstein and Lazar (n 6); Altman and Bland (n 6).
[28] Maddux and Rogers (n 13); Floyd, Prentice-Dunn and Rogers (n 13).
[29] Evidence Act 1872 (Bangladesh), s 65B.
[30] Evidence Act 1872 (Bangladesh), s 45; Evidence Act 1872 (Bangladesh), s 45A.

5.6 Integrated interpretation and sample boundaries

The study's distinctive result is not that the respondents simply trust or distrust AI. It is that support for some reliability conditions is widespread while support for a complete set is less common, and that the measured requirements do not clearly covary in the expected way after adjustment. Trust scholarship helps clarify why a statement about a desirable safeguard cannot be treated as willingness to accept vulnerability to an AI system. The measured pattern concerns the conditions respondents want around evidence, not an observed decision to rely on it.[31]

The sample composition further limits the reach of the argument. Students can provide useful information about legal education needs, but 253 student responses do not become professional evidence through aggregation. The 32-member professional subgroup gives an important descriptive check with wide uncertainty. Its results justify a larger, occupationally specific investigation rather than claims about Bangladesh's legal profession as a whole. The integrated interpretation is therefore a proposal for more explicit procedural education and better measurement, grounded in this survey pattern and existing forensic scholarship, rather than a diagnosis of national institutional failure.

6 Theoretical and Practical Implications

The findings matter because endorsement of a broad reliability requirement does not ensure endorsement of the procedures that make it examinable. Their implications concern measurement, education and evidence governance. Recommendations below combine the observed survey pattern with legal and forensic scholarship; they are not interventions whose effectiveness was tested in this cross-sectional study.

6.1 Theoretical implications

First, procedural reliability expectations should be disaggregated before they are modelled. Screening usefulness, independent validation, continuity, documentation and expert review have different objects. A single composite can conceal an important conjunction problem: many respondents endorse a salient safeguard while fewer affirm the entire set. The low internal consistency results reinforce caution, but the main theoretical point is semantic. Indicators must correspond to the construct being claimed. Future scale development should establish content coverage and factor structure before assigning a label such as perceived evidential reliability.[32]

Second, normative complementarity should not be confused with psychological integration. Forensic reasoning may require validation, custody and scrutiny together, but the respondent data do not show clearly that stronger validation agreement predicts the other requirements. This distinction

[31] Mayer, Davis and Schoorman (n 14); Jacovi and others (n 14).
[32] Cronbach (n 5); Sijtsma (n 5).

helps separate a governance model of what should be required from an empirical model of how people reason. Protection motivation can inform the latter only if threat and coping appraisals are measured directly and linked to a specified protective task. The present analysis supplies a bounded empirical problem for such work, rather than claiming a new validated psychological mechanism.[33]

Third, trust research can gain precision by locating the object of reliance. Trust in a developer, an expert, a particular tool and a case-specific inference may differ. A safeguard requirement is neither equivalent to any of these nor necessarily evidence of distrust. The conceptual contribution is to place desired reliability conditions alongside, but separately from, willingness to rely. Experimental research could then examine whether documented validation and accessible review change appropriately calibrated reliance on evidence of known quality.[34]

6.2 Educational and professional implications

Legal education should connect familiar custody and validation language with a concrete evidential file. An exercise could give students the original digital object, its handling history, an AI-generated classification and a short report. Students would then identify the claim being made, request missing method information and explain whether the output supports that claim. Assessment should reward recognition of uncertainty and limits, rather than merely the ability to repeat that AI requires validation. The present survey motivates this design; a separate evaluation would be needed to show whether it improves judgement.

For practitioners, a structured reporting checklist could record the task, material examined, acquisition history, tool and version, processing settings, validation basis, output, uncertainty and reviewer's reasoning. The purpose is to preserve a route from the evidence to the conclusion. Documentation should identify where a model assisted the process and where the expert made an independent inference. ISO and NIST guidance provide starting points for technical organisation, while local procedural obligations determine how the report is disclosed and used. A checklist should support professional judgement rather than substitute a completed form for demonstrated reliability.[35]

Review arrangements should identify who is competent to assess the particular task and

[33] Maddux and Rogers (n 13); Floyd, Prentice-Dunn and Rogers (n 13).
[34] Mayer, Davis and Schoorman (n 14); Jacovi and others (n 14); Durán and others (n 2).
[35] Kent and others (n 1); ISO/IEC 27042:2015 (n 4); Raji and others (n 11).

whether that person can question the output. Access to relevant records and authority to seek further examination are more informative than a generic statement that a human approved the result. Where resources are limited, institutions can prioritise preserving the information that later review would require. Such priorities should be assessed against the risks of the actual application; this survey supplies no cost-effectiveness estimate or evidence that a specific oversight model will succeed.[36]

6.3 Policy and societal implications

Policy guidance should distinguish admissibility conditions, technical validation and evidential weight in plain language. Training around sections 45, 45A and 65B can use examples showing how a digital record may meet a production requirement while the interpretation remains open to challenge. Guidance should also address confidentiality when legal material is processed through external AI systems. These recommendations aim to improve the intelligibility and contestability of evidence processes. They do not establish a new statutory duty or demonstrate that current institutional practices fail existing law.[37]

7 Conclusion Limitations and Future Research

This study examined procedural expectations surrounding AI assisted digital evidence analysis in a Bangladesh based mixed sample of 385 respondents. Independent validation and chain of custody received more endorsement than documentation and expert review. The paired differences across these two sets ranged from 16.1 to 22.9 percentage points, while only 29.4% endorsed all four safeguards. Adjusted models did not provide clear support for positive validation or custody associations with documentation or review. The findings therefore combine a clear descriptive disparity with uncertainty about conditional relationships.

The contribution is an item-specific account of how respondents connect complementary reliability safeguards. It avoids treating agreement about safeguards as evidence of actual AI accuracy, or a mixed sample as a national survey of professionals. The legal discussion clarifies why production and admissibility requirements should remain distinct from the scientific adequacy of an AI inference. Together, these boundaries make the evidence more useful: the results identify a focused question about procedural understanding without claiming to answer the separate questions of tool validation, courtroom admissibility or institutional compliance.

The cross-sectional design cannot establish temporal order or causal effects. Recruitment probabilities, dates and

[36] Green (n 12).
[37] Evidence Act 1872 (Bangladesh), s 45; Evidence Act 1872 (Bangladesh), s 45A; Evidence Act 1872 (Bangladesh), s 65B.

invitation counts are not available in the analytical record, limiting assessment of selection and nonresponse. Model-based intervals assume independent records; unknown clustering cannot be corrected without recruitment information. Students dominate the sample, and the professional subgroup is small and occupationally undifferentiated. Self-reported AI use does not establish forensic experience, and discordant experience fields introduce additional uncertainty.

Measurement also limits interpretation. The statements are broad, positively phrased and potentially sensitive to agreement tendencies. Their mixed content and low internal consistency do not support a validated reflective scale. Dichotomising endorsement simplifies presentation but combines neutral and negative responses; the full category counts and strict-threshold sensitivity analysis partially address this limitation. No actual evidence task, error rate, disclosure practice, court decision or observed review behaviour was measured. The analysis is exploratory and was not preregistered, so replication should be treated as an important next step.

Future surveys should recruit defined occupational groups and record recruitment channels, invitation counts and relevant forensic experience. Cognitive interviews could test whether participants interpret validation, custody and documentation consistently. A revised instrument should distinguish understanding of safeguards, endorsement of safeguards, institutional capacity and willingness to rely on a specified output. Those dimensions can then be evaluated through content validation and an appropriate measurement model rather than assumed from section labels.

Experiments should vary evidential quality and disclosure separately. Participants could assess outputs with known error characteristics, differing custody records, access to validation material and opportunities for expert questioning. Such designs would reveal whether additional information improves calibration rather than simply increasing confidence. Longitudinal educational studies could test whether task-based instruction closes the observed endorsement differences and improves performance. Comparative legal research could then examine how these procedures interact with different admissibility rules without assuming that a governance framework developed elsewhere automatically governs Bangladesh. The present findings provide a focused starting point for that programme: reliability conditions need to be connected explicitly, and their practical value must be demonstrated through evidence rather than inferred from favourable attitudes.

Declarations

Ethical approval and oversight

Formal ethical approval was not obtained because, at the time of the study, the authors' institutional setting did not provide an accessible formal review mechanism applicable to this non-interventional adult questionnaire research. The study nevertheless followed participant-protection safeguards, including informed consent, voluntary participation, the opportunity to discontinue, an anonymous questionnaire

and aggregate reporting. The methodological safeguards are detailed in Section 3.8. The absence of an accessible review mechanism describes the institutional context; responsibility for ethical conduct remained with the researchers.

Informed consent

Informed consent was obtained from adult participants before questionnaire completion. The questionnaire explained the study purpose, voluntary participation, anonymity and the option to discontinue. All analysed records contain affirmative consent and adult-eligibility responses.

Data availability

The numerical data supporting the reported analyses are held by the research team. Requests for de-identified data and analytical specifications may be directed to the corresponding author. Any sharing is subject to participant confidentiality and applicable institutional requirements. The contextual-text field was excluded from the analysis.

Author contributions

Mst. Dilara Islam Anny: Lead student researcher; conceptualisation; preparation of study materials; organisation and curation of survey data; manuscript development. The project was conducted under the supervision of Prof. Dr Kazi Abdul Mannan.

Kazi Abdul Mannan: Conceptual and methodological guidance; supervision; interpretation of statistical findings; critical review and editing of the manuscript.

Generative AI disclosure

Generative AI assisted manuscript drafting, language editing, organisation and the development of analytical code. Statistical estimates were computed programmatically from the numerical survey data. Responsibility for the study's analysis, interpretation and reporting rests with the authors.

Bibliography

Altman DG and Bland JM, ‘Absence of Evidence Is Not Evidence of Absence’ (1995) 311 BMJ 485

Cochran WG, *Sampling Techniques* (3rd edn, John Wiley & Sons 1977)

Cronbach LJ, ‘Coefficient Alpha and the Internal Structure of Tests’ (1951) 16 Psychometrika 297

Durán JM and others, ‘From Understanding to Justifying: Computational Reliabilism for AI-Based Forensic Evidence Evaluation’ (2024) 9 Forensic Science International: Synergy 100554

Efron B and Tibshirani RJ, *An Introduction to the Bootstrap* (Chapman & Hall 1993)

Floyd DL, Prentice-Dunn S and Rogers RW, ‘A Meta-Analysis of Research on Protection Motivation Theory’ (2000) 30 Journal of Applied Social Psychology 407

Green B, ‘The Flaws of Policies Requiring Human Oversight of Government Algorithms’ (2022) 45 Computer Law & Security Review 105681

Holm S, ‘A Simple Sequentially Rejective Multiple Test Procedure’ (1979) 6 Scandinavian Journal of Statistics 65

International Organization for Standardization and International Electrotechnical Commission, *Information Technology—Security Techniques—Guidelines for Identification, Collection, Acquisition and Preservation of Digital Evidence* (ISO/IEC 27037:2012, ISO 2012)

——, *Information Technology—Security Techniques—Guidelines for the Analysis and Interpretation of Digital Evidence* (ISO/IEC 27042:2015, ISO 2015)

Jacovi A and others, 'Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI' in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2021) 624

Kent K and others, *Guide to Integrating Forensic Techniques into Incident Response* (NIST Special Publication 800-86, NIST 2006)

Lyle JR and others, *Digital Investigation Techniques: A NIST Scientific Foundation Review* (NISTIR 8354, NIST 2022)

Maddux JE and Rogers RW, 'Protection Motivation and Self-Efficacy: A Revised Theory of Fear Appeals and Attitude Change' (1983) 19 Journal of Experimental Social Psychology 469

Mayer RC, Davis JH and Schoorman FD, 'An Integrative Model of Organizational Trust' (1995) 20 Academy of Management Review 709

McNemar Q, 'Note on the Sampling Error of the Difference between Correlated Proportions or Percentages' (1947) 12 Psychometrika 153

Mitchell M and others, 'Model Cards for Model Reporting' in *Proceedings of the Conference on Fairness, Accountability, and Transparency* (ACM 2019) 220

Raji ID and others, 'Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing' in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (ACM 2020) 33

Sijtsma K, 'On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha' (2009) 74 Psychometrika 107

Stefanski LA and Boos DD, 'The Calculus of M-Estimation' (2002) 56 American Statistician 29

Stoykova R, 'A New Right to Procedural Accuracy: A Governance Model for Digital Evidence in Criminal Proceedings' (2024) 55 Computer Law & Security Review 106040

Wasserstein RL and Lazar NA, 'The ASA Statement on p-Values: Context, Process, and Purpose' (2016) 70 American Statistician 129

Wilson EB, 'Probable Inference, the Law of Succession, and Statistical Inference' (1927) 22 Journal of the American Statistical Association 209

Appendix A Questionnaire statements

Table A1 reproduces the focal digital evidence statements and the verification adjustment. Retaining exact wording makes the distinction between perceived capability and normative requirements inspectable.

Table A1. Exact focal and verification statements

Code	Statement
DER1	AI tools can assist in organizing or screening large volumes of digital evidence.
DER2	AI-assisted evidence analysis should be independently validated before legal reliance.
DER3	A clear chain of custody remains essential when AI tools are used.
DER4	Methods used by AI tools should be sufficiently documented for scrutiny.
DER5	Human expert review is necessary for high-stakes interpretation of AI-assisted digital evidence.
AIL2	I verify important AI-generated information before relying on it.

Source: Group 6 Codebook. Scale: one strongly disagree, two disagree, three neutral, four agree and five strongly agree.

The response scale is common across the listed items. No item directly asks whether an identified AI system achieved an acceptable accuracy level in a specified legal case.

Appendix B Reproducibility specifications

The primary dataset contains 385 records. Endorsement is coded one for scores four or five and zero otherwise. For DER2-DER5, every unordered pair is compared using an exact two-sided binomial test on discordant pairs; Holm adjustment covers all six p values. Bootstrap intervals use 10,000 whole-record resamples, seed 20261025, and the 2.5th and 97.5th percentiles of each paired difference. Wilson intervals use the standard normal 0.975 quantile.

For each of DER4 and DER5, logistic regression includes an intercept and DER2–3, DER3–3, DER1–3, AIL2–3, any formal AI training, current advocate or legal professional status, frequency–2 and age 27+. Frequency is zero never, one less than monthly, two monthly, three weekly and four daily or almost daily. Training is zero only for No; all other recorded training categories are one. Professional status is one only for the matching current-status category. Age 27+ combines the 27-30 and 31+ groups. HC1 covariance equals the sandwich estimate multiplied by n/(n–k). Reported odds ratios exponentiate coefficients; intervals exponentiate coefficient ±1.96 standard errors.

The consistent-use sensitivity excludes records with Never in either experience or frequency. The law-student sensitivity retains current law students and removes the invariant professional-status indicator. The strict outcome equals one only for score five. No contextual-text content enters any calculation. The numerical specifications are available from the corresponding author

under the data-sharing terms in Section 8.

Appendix C Complete adjusted coefficients

Table C1 reports all non-intercept terms so that the selected focal associations can be assessed in the context of the full specification. Contextual coefficients are exploratory adjustments and are not presented as independent confirmatory discoveries.

Table C1. Complete adjusted non-intercept coefficients

Predictor	Documentation OR (95% CI)	Review OR (95% CI)
DER2 independent validation	1.06 (0.80-1.41)	0.94 (0.70-1.28)
DER3 chain of custody	1.14 (0.85-1.54)	1.17 (0.87-1.58)
DER1 screening usefulness	1.12 (0.87-1.45)	0.80 (0.62-1.04)
AIL2 self-reported verification	1.30 (0.97-1.74)	1.26 (0.92-1.71)
Any formal AI training	1.28 (0.83-1.96)	1.54 (0.99-2.39)
Advocate/legal professional	1.05 (0.47-2.32)	0.63 (0.29-1.39)
AI frequency category	1.22 (1.00-1.50)	1.12 (0.91-1.37)
Age 27+	1.30 (0.76-2.23)	1.79 (1.03-3.10)

N = 385. HC1 pointwise intervals. Training compares any training with none; professional status compares advocates/legal professionals with others; age compares 27+ with younger groups. Frequency is per category. See Appendix B for all coding rules.

Intervals are pointwise. The professional-status coefficient in the full sample is an adjustment based on 32 professional records, not a well-powered occupational comparison or a professional-only estimate.

Table of legislation

Bangladesh: Evidence Act 1872, ss 45, 45A and 65B (as amended by the Evidence (Amendment) Act 2022).