Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explanation and Validation as Conditions for Trust in Judicial AI in Bangladesh

Dr Kazi Abdul Mannan

drkaziabdulmannan@gmail.com

Center for Academic & Professional Career Development and Research (CAPCDR)
https://orcid.org/0000-0002-7123-132X

Maksuda Ferdous

Shanto-Mariam University of Creative Technology https://orcid.org/0009-0002-2810-8665

Research Article

Keywords: explainable artificial intelligence, conditional trust, judicial decision support, procedural justice, independent validation, Bangladesh, SDG 16

Posted Date: October 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11523361/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Explanation and Validation as Conditions for Trust in Judicial AI in Bangladesh

Maksuda Ferdous

Department of Law, Shanto-Mariam University of Creative Technology, Uttara, Dhaka, Bangladesh

Email: maksudaferdous883@gmail.com

ORCID: https://orcid.org/0009-0002-2810-8665

Prof. Dr Kazi Abdul Mannan

Department of Business Administration, Faculty of Business, Shanto-Mariam University of Creative Technology, Dhaka, Bangladesh

Email: drkaziabdulmannan@gmail.com

ORCID: https://orcid.org/0000-0002-7123-132X

Abstract

Judicial artificial intelligence requires grounds for warranted reliance rather than confidence alone. This study examines how respondents in Bangladesh distinguish understandable reasons, independent validation and opportunities to challenge AI-supported recommendations. A cross-sectional analysis of 385 adult, AI-aware respondents uses seven questionnaire items without treating their mixed normative and conditional wording as a validated explainability-trust scale. The analysis combines paired endorsement comparisons, proportional-odds models with robust covariance, multiplicity adjustment and paired-outcome coefficient comparisons. Independent-validation-contingent trust was endorsed by 247 respondents (64.2%), compared with 215 (55.8%) for explanation-contingent trust: a paired difference of 8.3 percentage points (95% bootstrap interval 1.6-15.1; Holm-adjusted $p = .021$). Access to review or appeal received greater endorsement than the ability to question recommendations (85.2% versus 65.2%; difference 20.0 points, 95% interval 14.3-25.7). None of the eight focal associations between four governance expectations and the two trust propositions remained distinguishable from zero after Holm adjustment. The joint test of differences between the four outcome-specific coefficients was also inconclusive ($p = .236$). These findings identify a descriptive separation between explanations, validation and remedies while providing limited adjusted evidence about how those expectations align within respondents. The predominantly student sample does not establish judicial practice, actual reliance or causal effects of explanations. The article develops a layered account of judicial AI governance that connects understandable reasons with independent evaluation and effective review, and specifies a measurement and experimental agenda for testing calibrated reliance.

Keywords: explainable artificial intelligence; conditional trust; judicial decision support; procedural justice; independent validation; Bangladesh; SDG 16

1. Introduction

Artificial intelligence can assist legal work by organising documents, retrieving information and identifying patterns. These capabilities become institutionally sensitive when an output enters a process affecting rights, obligations or access to justice. A useful recommendation does not automatically supply a defensible reason for a judicial decision. The relevant question is how a human decision-maker can examine the recommendation, identify its limitations and remain accountable for its use. Surden's account of AI and law provides a starting point by distinguishing practical computational capabilities from broader claims about machine judgement. This distinction is especially valuable when survey respondents have experience of general-purpose AI rather than a specified judicial application.[1]

The present article addresses trust under those conditions. It examines what respondents say would make an AI recommendation more trustworthy, together with their expectations of disclosure, challengeability and review. Its object is therefore a set of prospective conditions for trust. It does not measure confidence in a particular court, test a deployed decision-support system or assess the legal quality of an AI-generated judgment. Keeping these objects separate makes it possible to draw useful conclusions from the questionnaire without turning a survey of expectations into evidence of institutional performance.

International guidance increasingly places judicial AI within a framework of human responsibility. UNESCO's 2025 guidelines provide an ethical and operational reference for courts and tribunals; the CEPEJ charter links judicial AI to fundamental rights, non-discrimination, transparency and user control. These instruments provide comparative governance benchmarks. They do not, by their existence, establish binding domestic duties in Bangladesh. Their significance here is that they locate explanations within a wider institutional arrangement rather than presenting transparency as a complete solution.[2]

1.1 Research problem and context

The research problem arises from a frequent compression of several ideas into the phrase explainability builds trust. An understandable account of a recommendation, a disclosure of influential factors, independent validation and a route to appeal perform different functions. A person may want all four without trusting an unreliable system. Conversely, a plausible explanation can encourage reliance even when an output is incorrect. Trust in automation scholarship therefore distinguishes confidence from appropriate reliance. The desirable outcome is a relationship between reliance and warranted capability, not the maximisation of agreement with favourable statements about technology.[3]

---

[1] Harry Surden, 'Artificial Intelligence and Law: An Overview' (2019) 35 Georgia State University Law Review 1305.

[2] Juan David Gutiérrez, *Guidelines for the Use of AI Systems in Courts and Tribunals* (UNESCO 2025); European Commission for the Efficiency of Justice, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment* (Council of Europe 2018).

[3] John D Lee and Katrina A See, 'Trust in Automation: Designing for Appropriate Reliance' (2004) 46 Human Factors 50; Kevin Anthony Hoff and Masooda Bashir, 'Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust' (2015) 57 Human Factors 407.

Bangladesh provides the setting for the supplied Group 3 survey. The official national AI policy portal identifies the 2026-2030 policy as a draft, including a version 2.0 consultation document. This policy process indicates governance attention, but the draft is not treated here as enacted legislation. The article contributes evidence about respondents' expectations within this context, without claiming that the sample represents the judiciary or that a particular judicial AI programme has been implemented. The legal-policy discussion consequently remains a set of institutionally relevant recommendations rather than a compliance assessment.[4]

The sample also requires a precise description of the people whose responses are being analysed. Of 385 records, 250 identify the respondent as a law student and 36 as an advocate or legal professional. Students are relevant to future legal practice and legal education, but their responses cannot substitute for evidence from judges, litigants or experienced practitioners. This composition is part of the study's substantive meaning. It directs the findings towards early governance expectations and identifies professional validation as a next research step.

1.2 Gap and distinctive contribution

Three analytical gaps organise the contribution. First, explanation and trust are often combined at the measurement stage, making a subsequent explanation-to-trust regression conceptually circular. Second, aggregate agreement can conceal different response patterns across closely related governance propositions. Third, a predictor may appear stronger for one outcome than another simply because separate models produce different p-values. Establishing a difference requires a direct comparison that respects the correlation of the outcomes. The present analysis addresses all three problems through item-level interpretation, paired comparisons and a covariance-aware comparison of model coefficients.

The article develops a focused contribution to the study of judicial AI by comparing two explicitly conditional trust statements: greater trust when reasons are understood and greater trust following independent validation. It then examines whether disclosure, challengeability, non-reliance on unexplained outputs and access to review align with these propositions. The distinctive contribution is an analysis of the conditions respondents articulate, together with a clear account of what the instrument cannot identify.

1.3 Objectives and research questions

The first objective is to describe the full response distributions for the two trust propositions and five related governance or verification items. The second is to estimate paired differences between validation and explanation, and between review or appeal and challengeability. The third is to examine adjusted associations between four governance expectations and each trust proposition. The fourth is to compare those associations across the two outcomes without interpreting differences in statistical significance as evidence of different effects.

The resulting questions are: how frequently are the two conditions for trust endorsed; how do their endorsements differ within respondents; which governance expectations align with each

[4] Information and Communication Technology Division, *National Artificial Intelligence Policy 2026–2030* (Draft Version 2.0, 2026) <https://aipolicy.gov.bd/docs/national-ai-policy-bangladesh-2026-2030-draft-v2.0.pdf> accessed 26 October 2026.

condition after adjustment; and is there evidence that those associations differ across conditions? The answers support a layered account of judicial AI governance while retaining an exploratory statistical interpretation. Section 2 develops the theoretical foundation, Section 3 describes measurement and analysis, Section 4 reports results, and Sections 5-7 develop their interpretation, implications and conclusions.

2. Literature Review and Theoretical Foundation

The theoretical foundation connects explainable AI, trust in automation and procedural justice. Each addresses a different part of the problem: explanations concern the intelligibility of a recommendation, trust concerns willingness to rely under uncertainty, and procedural justice concerns the organisation and evaluation of decision-making relationships. The analysis draws on these traditions without claiming to estimate a comprehensive latent model of any one of them. The purpose of the review is to establish distinctions that the available questionnaire can meaningfully examine and to prevent stronger causal or institutional interpretations from being attached to its responses.

2.1 Explanation as a communicative relationship

Explainable AI includes methods for rendering aspects of model behaviour understandable to human audiences. Technical taxonomies distinguish approaches and purposes, but a technical output becomes useful only in relation to an audience and a task. A feature ranking, for example, can identify statistical influence without supplying a legally sufficient account of relevance, reliability or justification. The distinction matters for judicial decision support because the audience may include judges, advocates, parties and reviewing institutions, each with different informational needs. The present questionnaire records expectations about explanations; it does not compare technical explanation methods or test whether respondents understand them.[5]

Miller's review directs attention to explanation as a social process shaped by the questions people ask and the alternatives they consider. Gregor and Benbasat similarly locate explanations within the use of intelligent systems rather than treating their availability as automatically beneficial. These perspectives support a relational interpretation of EXT1: the statement links increased trust to understanding the main reasons behind a recommendation. It does not establish that an explanation was provided, that understanding occurred or that the reasons were accurate. Those distinctions should remain visible when interpreting an apparently straightforward agreement response.[6]

The legal setting adds a distinction between describing a computational process and justifying the use of its output. A model might explain why it assigned a score while relying on variables that a decision-maker should not use. It might faithfully reproduce a flawed data-generating process. It might also offer an accessible explanation that omits uncertainty or important

[5] Alejandro Barredo Arrieta and others, 'Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI' (2020) 58 Information Fusion 82.

[6] Tim Miller, 'Explanation in Artificial Intelligence: Insights from the Social Sciences' (2019) 267 Artificial Intelligence 1; Shirley Gregor and Izak Benbasat, 'Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice' (1999) 23 MIS Quarterly 497.

exceptions. These possibilities make explanation an entry point for scrutiny rather than a substitute for scrutiny. Research on legal hallucinations further illustrates why fluent output cannot by itself establish legal reliability, although that research concerns tested systems and tasks rather than the respondents in this study.[7]

2.2 Trust, trustworthiness and calibrated reliance

Organisational trust scholarship separates the trustor's willingness to accept vulnerability from the characteristics of the trustee. Mayer, Davis and Schoorman identify dimensions of perceived trustworthiness, while Rousseau and colleagues emphasise the common role of vulnerability and positive expectations across disciplines. These accounts caution against treating an agreement score as a direct measure of an AI system's objective trustworthiness. A respondent's view that validation would increase trust concerns an anticipated basis for confidence, not an observation that a specific system deserves confidence.[8]

In automation research, appropriate reliance depends on the interaction between user expectations, system capability and context. Lee and See, Hoff and Bashir, and Glikson and Woolley provide complementary foundations for examining this interaction. For present purposes, the central implication is that trust should be calibrated to evidence and limitations. A successful explanation may sometimes reduce trust by revealing an unsuitable basis for a recommendation. A validation exercise may also appropriately prevent deployment. Thus, a governance arrangement should remain effective even when its immediate outcome is less reliance.[9]

EXT1 and EXT4 capture two different grounds respondents say would increase trust. The former concerns their understanding of reasons; the latter concerns independent validation. Neither asks whether they would rely on a wrong recommendation, whether they would recognise an error or how they would change their behaviour after encountering contradictory evidence. Jacovi and colleagues' treatment of trust in AI helps clarify why these omissions matter. The article accordingly uses the terms explanation-contingent and validation-contingent trust propositions as transparent descriptions of item content, not as names for newly validated scales.[10]

2.3 Transparency, validation and institutional accountability

Transparency provides access to information, but accountability requires more than visibility. Ananny and Crawford identify limits in equating disclosure with effective knowledge of

[7] Joshua A Kroll and others, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 633; Matthew Dahl and others, 'Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models' (2024) 16 Journal of Legal Analysis 64.

[8] Roger C Mayer, James H Davis and F David Schoorman, 'An Integrative Model of Organizational Trust' (1995) 20 Academy of Management Review 709; Denise M Rousseau and others, 'Not So Different After All: A Cross-Discipline View of Trust' (1998) 23 Academy of Management Review 393.

[9] Lee and See (n 3); Hoff and Bashir (n 3); Ella Glikson and Anita Williams Woolley, 'Human Trust in Artificial Intelligence: Review of Empirical Research' (2020) 14 Academy of Management Annals 627.

[10] 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.

complex systems. Bovens' account of accountability instead foregrounds a relationship in which an actor must explain and justify conduct to a forum capable of questioning and evaluating it. Applied to judicial AI, disclosure becomes institutionally useful when someone can assess its significance, request further information and act on identified problems. This is a conceptual application rather than a finding about existing Bangladeshi institutions.[11]

Independent validation occupies a different position within that relationship. It can examine whether a system performs adequately for a specified task, population and setting, rather than relying on the persuasiveness of an individual output. Algorithmic auditing research and public-administration scholarship describe organisational responsibilities that extend across development, procurement and use. The survey does not establish what respondents mean by independence or validation, so EXT4 cannot identify preferences for a particular evaluator or testing regime. It nevertheless allows a descriptive comparison with understanding reasons, provided that the two statements are not represented as equivalent interventions.[12]

The distinction also affects interpretation of the four governance predictors. EXT2 concerns disclosure of influential factors; EXT3 concerns the ability to question or challenge recommendations; EXT5 expresses a restriction on reliance when recommendations are unexplained; and ABPF3 concerns access to review or appeal for affected people. Each involves an institutional expectation. Their content can overlap without being interchangeable. The article therefore estimates their separate associations while avoiding an undifferentiated transparency score that would conceal the different relationships between information, action and remedy.

2.4 Procedural justice and reviewability

Procedural justice provides a complementary lens because the acceptability of a decision-making process is not exhausted by its outcome. Tyler's account helps explain the importance of fair treatment and the opportunity to be heard. The questionnaire's challengeability and review items touch this domain, but they do not cover the full procedural justice construct. They say little about perceived neutrality, respectful treatment or the credibility of institutional motives. Accordingly, procedural justice informs the interpretation of the items; it is not declared empirically validated by their inclusion.[13]

Reviewability also differs from an opportunity to question a recommendation before a decision. Cobbe, Lee and Singh's framework directs attention to the wider decision-making process, including the information needed for subsequent assessment. A party may possess a formal review route but lack access to the material necessary to use it. A user may be able to question a screen output without having any authority to change the institutional decision. These

[11] Mike Ananny and Kate Crawford, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 New Media & Society 973; Mark Bovens, 'Analysing and Assessing Accountability: A Conceptual Framework' (2007) 13 European Law Journal 447.

[12] 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; Madalina Busuioc, 'Accountable Artificial Intelligence: Holding Algorithms to Account' (2021) 81 Public Administration Review 825.

[13] Tom R Tyler, *Why People Obey the Law* (Yale University Press 1990).

distinctions help explain why ABPF3 and EXT3 should be compared as related but non-identical propositions rather than scored as interchangeable indicators.[14]

Human oversight is similarly a design problem rather than a sufficient label. Green's critique of oversight requirements and Alon-Barkat and Busuioc's work on human-AI interaction caution against assuming that a nominal human decision-maker will necessarily correct algorithmic error. The present survey cannot test automation bias or selective adherence. Their relevance is theoretical: a practical response to respondents' expectations must specify the information, time, authority and review arrangements through which a person can assess an AI contribution. Merely retaining a signature at the end of the process would leave those questions unanswered.[15]

2.5 Acceptance models and measurement boundaries

The Technology Acceptance Model explains the role of perceived usefulness and ease of use in acceptance. It remains relevant as a contextual account of why decision-support tools might attract interest. However, the present trust items do not constitute the model's standard measures, and the study does not estimate a TAM mediation pathway. Extending a familiar theory by relabelling heterogeneous questionnaire items would create an appearance of theoretical precision without a corresponding measurement basis. TAM therefore supplies a supporting perspective rather than the article's statistical architecture.[16]

Scale-development research requires attention to content, dimensionality and validation. It is especially relevant here because the EXT block mixes conditional trust, normative disclosure, challengeability and a rule against unexplained reliance. Treating these statements as reflective manifestations of a single latent trait would require justification beyond a reliability coefficient. Similarly, dividing the block into an explainability predictor and trust outcome after observing the data could embed semantic overlap in the purported structural relationship. The present analysis avoids that problem by retaining the original item meanings and modelling two conditional outcomes separately.[17]

2.6 Exploratory propositions and analytical framework

Four exploratory propositions guide the adjusted analysis. Greater endorsement of factor disclosure (H1), challengeability (H2), non-reliance on unexplained outputs (H3) and access to review or appeal (H4) may each align positively with the two conditional trust propositions.

---

[14] Jennifer Cobbe, Michelle Seng Ah Lee and Jatinder Singh, 'Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems' in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2021) 598.

[15] Ben Green, 'The Flaws of Policies Requiring Human Oversight of Government Algorithms' (2022) 45 Computer Law & Security Review 105681; Saar Alon-Barkat and Madalina Busuioc, 'Human–AI Interactions in Public Sector Decision Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice' (2023) 33 Journal of Public Administration Research and Theory 153.

[16] Fred D Davis, 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology' (1989) 13 MIS Quarterly 319.

[17] Godfred O Boateng and others, 'Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer' (2018) 6 Frontiers in Public Health 149; Adamantios Diamantopoulos and Heidi M Winklhofer, 'Index Construction with Formative Indicators: An Alternative to Scale Development' (2001) 38 Journal of Marketing Research 269.

These are response-association propositions, developed for the current analysis rather than preregistered causal hypotheses. Each yields one coefficient for EXT1 and one for EXT4, creating a family of eight focal comparisons. The framework also asks whether the four associations differ across outcomes, using a direct paired-outcome comparison.

Figure 1 shows the framework. The four governance expectations enter both models alongside reported verification practice, training, use frequency and age. The diagram deliberately separates the two outcomes and represents associational paths only. A positive coefficient would indicate alignment among reported expectations after adjustment; it would not show that introducing a disclosure or appeal mechanism increases trust. Conversely, an imprecise estimate would not demonstrate that the corresponding safeguard lacks legal or institutional value.

Figure 1. Item-level framework for conditional trust in judicial AI. Arrows represent exploratory associations, not causal effects.

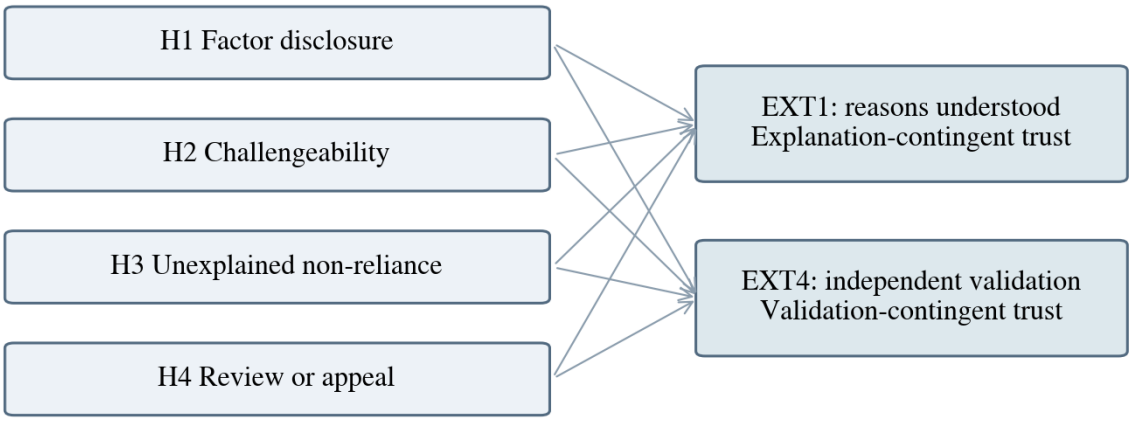


Figure 1. Item-level framework for conditional trust in judicial AI

The framework generates two forms of contribution. Paired descriptive contrasts identify differences in what respondents endorse, while adjusted models examine the extent to which endorsements covary. These questions can produce different answers because a commonly endorsed safeguard need not distinguish respondents who more strongly endorse a conditional trust statement. Keeping the questions separate is essential to interpreting the findings that follow.

3. Methodology

This study uses the supplied Group 3 database to examine prospective conditions for trust in judicial AI. The design is cross-sectional and observational. All analyses are based on numerical questionnaire responses, with the optional contextual-text field excluded. The

methodological approach prioritises traceable item interpretation, complete accounting of the 385 records and explicit treatment of paired responses. It combines descriptive estimates with exploratory adjusted analysis; it does not infer causal effects or claim that questionnaire agreement demonstrates the performance of a judicial AI system.

3.1 Data source and study population

The analytical source is the Data sheet of Group-3-Law-SDG-Database, Summer 2026, interpreted alongside its README and Codebook. The research team confirmed that the numerical responses originated from a collected survey; synthetic labels attached to the identifier description and optional comment field were not treated as a description of the numerical responses. Respondent_ID served only to check record uniqueness. The contextual-text field was not analysed. The manuscript's empirical claims concern the numerical survey records, whose source and collection documentation remain the research team's responsibility.

The database contains 385 records, all with affirmative consent, adult eligibility and relevant AI-awareness responses. These establish the recorded eligibility criteria, not independent evidence of participants' professional competence or experience of judicial systems. No inference of a judges-only sample is made. The supplied materials do not document a probability sampling frame, invitations, response denominator or fieldwork dates. The sample is therefore analysed without design weights, a response-rate estimate or a national representativeness claim. Its size is the observed number of eligible records rather than a reconstructed a priori power calculation.

3.2 Variables and measurement

The primary outcomes are EXT1 and EXT4. EXT1 states that trust is greater when the respondent understands the main reasons behind an AI recommendation; EXT4 states that independent validation would increase trust in decision-support tools. The focal predictors are EXT2, EXT3, EXT5 and ABPF3. AIL2 measures reported verification of important AI-generated information. All use five response categories from strongly disagree to strongly agree. Exact wording appears in Appendix A, and Table 1 specifies the analytical roles before results are presented.

Table 1. Item content and analytical roles

Code	Abbreviated content	Role
EXT1	Trust increases with understood reasons	Outcome 1
EXT4	Trust increases with independent validation	Outcome 2
EXT2	Disclosure of influential factors	Focal H1
EXT3	Ability to question or challenge	Focal H2
EXT5	Non-reliance on unexplained outputs	Focal H3
ABPF3	Access to review or appeal	Focal H4
AIL2	Verification of important AI information	Adjustment

Source: Group 3 Codebook. Appendix A gives exact wording. All items are retained separately; none is presented as a validated multi-item construct.

These roles follow the wording rather than the original block labels. No reverse coding is applied: higher EXT5 indicates stronger endorsement of non-reliance on unexplained recommendations, not a general negative attitude towards AI. The analysis does not create a total EXT score or report latent validity indices for a scale that has not been established. Means describe the numerical coding; distributions and ordinal models retain the substantive importance of the response categories.

3.3 Data integrity and covariates

The audit checked eligibility fields, missingness, permitted values, duplicate identifiers, repeated complete item vectors and uniform responses across the 20 agreement items. No focal item was missing, so all 385 records entered the primary analysis. Twelve tool-name entries were missing, but that field was not used. Twenty-four records reported inconsistent combinations of never-use experience and non-never frequency, or the converse. These were retained because the inconsistency did not invalidate their completed focal responses; a restricted-use sensitivity analysis excluded both groups.

The models include five agreement-point variables: the four focal expectations and verification practice. Agreement predictors are centred at neutral and enter linearly. Contextual covariates comprise any AI training versus none; AI-use frequency coded zero through four from never to daily or almost daily; and age 27 or above versus younger. These parsimonious codings limit parameter count but impose linearity for frequency and compress age information. They are adjustment choices rather than identified confounder controls. A sensitivity specification adds law-student status because of the sample’s composition.

3.4 Descriptive and paired analyses

For each analytical item, the analysis reports category counts, mean and sample standard deviation, endorsement at categories four or five, and pointwise 95% Wilson intervals. These intervals quantify uncertainty under an independent-record model; they do not repair non-

probability recruitment. Two paired endorsement contrasts were specified for the analysis: EXT4 minus EXT1 and ABPF3 minus EXT3. The former compares trust conditions; the latter compares two procedural expectations with different beneficiaries and institutional meanings. Neither contrast estimates the effect of changing one feature into another.[18]

Exact McNemar tests use only discordant response pairs. Holm adjustment covers the two comparisons. Percentage-point differences receive percentile intervals from 10,000 whole-record bootstrap resamples using seed 20261026. Resampling whole records preserves the pairing. Joint tables additionally show respondents endorsing both, one or neither proposition. Non-endorsement includes neutrality and should not be interpreted as opposition.[19]

3.5 Ordinal models and diagnostic assessment

Proportional-odds logistic models estimate the association of each predictor with higher responses on EXT1 and EXT4. Each outcome has a single category-one response; categories one and two are therefore pooled as disagreement for modelling, producing four ordered categories. The original five-category distributions remain fully reported. The model is logit $\Pr(Y \leq k \mid X) = \alpha k - X\beta$, so $\exp(\beta)$ is the odds ratio for a higher response per unit increase in a predictor. There are eight slopes and three thresholds per model.[20]

Maximum likelihood estimation uses analytic scores, a numerical observed Hessian and HC1 sandwich covariance with correction $N/(N - 11)$. A second optimiser checks the fitted coefficients. A cumulative-logit common-slope diagnostic compares slopes at the three cumulative cutpoints, retaining within-record dependence among those fits; its Wald test has 16 degrees of freedom. The stacked cumulative fits use record-level sandwich covariance. No diagnostic rejection was found, so the ordinal models remain primary.[21]

3.6 Multiplicity and outcome comparisons

Holm adjustment covers the eight focal ordinal coefficients, comprising four governance expectations across two outcomes. Covariate intervals and model-fit summaries are descriptive. To compare outcome-specific slopes, the analysis subtracts the EXT1 coefficients from the EXT4 coefficients and constructs covariance from differences between their per-record influence vectors. This retains the correlation generated by fitting both outcomes to the same respondents. The four differences receive a joint Wald test and separate Holm-adjusted comparisons. Exponentiated differences are ratios of ordinal odds ratios, not changes in endorsement percentages.

---

[18] Edwin B Wilson, 'Probable Inference, the Law of Succession, and Statistical Inference' (1927) 22 Journal of the American Statistical Association 209.

[19] Quinn McNemar, 'Note on the Sampling Error of the Difference between Correlated Proportions or Percentages' (1947) 12 Psychometrika 153; 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).

[20] Peter McCullagh, 'Regression Models for Ordinal Data' (1980) 42 Journal of the Royal Statistical Society Series B 109.

[21] Leonard A Stefanski and Dennis D Boos, 'The Calculus of M-Estimation' (2002) 56 American Statistician 29; Rollin Brant, 'Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression' (1990) 46 Biometrics 1171.

This approach avoids concluding that two associations differ because one p-value crosses a conventional threshold and another does not. The joint comparison is exploratory, and its uncertainty remains conditional on the specification and independent records. The study reports estimates and intervals alongside adjusted p-values, retaining the distinction between limited evidence and evidence of no relationship.[22]

3.7 Sensitivity and reproducibility

Two sensitivity analyses refit the ordinal models after adding student status and after excluding the 24 inconsistent-use records. Binary logistic models additionally compare endorsement versus lower categories and strong agreement versus all lower categories. These thresholds change the estimand and are reported as descriptive sensitivity checks, not additional confirmatory discoveries. Binary coefficients were cross-checked using an independent logistic-regression implementation. The primary and sensitivity models use no imputation, automatic stepwise selection or outcome-derived composite score.

The analytical sequence can be reproduced from the named Data and Codebook sheets: check eligibility; inspect values; define outcomes and covariates; generate paired indicators; estimate the models; compute covariance and multiplicity adjustments; and tabulate results. Reported intervals are pointwise, except that p-values are adjusted within their stated families. No institutional or recruitment-cluster identifiers are available; robust individual covariance cannot account for unrecorded clustering. Analytical specifications may be requested under the data-availability statement.

3.8 Ethics and participant protection

Ethical approval for this study was obtained from the Center for Academic & Professional Career Development and Research (CAPCDR), an external institutional body, under approval reference CAPCDR: SMUCT/LLB/SS-20261012-02. External ethical review was sought because Shanto-Mariam University of Creative Technology does not have an institutional ethics committee responsible for reviewing this type of research.

Safeguards reported by the research team comprised an explanation of purpose, informed consent before completion, voluntary participation, the option to discontinue, an anonymous questionnaire and aggregate reporting. All analysed records contain affirmative adult-eligibility and consent responses. The analysis excludes contextual text, reports no identifiable narratives and avoids releasing record-level data in the manuscript. Any subsequent sharing must preserve confidentiality and comply with applicable requirements. The ethical account rests on the reported participant-protection procedures and the researchers' responsibility for their implementation.

[22] Ronald L Wasserstein and Nicole A Lazar, 'The ASA Statement on p-Values: Context, Process, and Purpose' (2016) 70 American Statistician 129; Douglas G Altman and J Martin Bland, 'Absence of Evidence Is Not Evidence of Absence' (1995) 311 BMJ 485.

4 Results

The results distinguish how often a proposition is endorsed from whether it aligns with other expectations after adjustment. Tables 2-4 establish the sample and response distributions; Tables 5-6 retain within-respondent comparisons; and Tables 7-10 report models, direct outcome comparisons and sensitivity estimates. All primary analyses use 385 records. Percentages use this denominator unless another is specified. The findings concern questionnaire responses about prospective judicial AI conditions and should not be interpreted as observations of judicial decisions or actual changes in trust.

4.1 Respondent profile and analytical completeness

Table 2 reports the complete profile rather than describing all respondents as legal professionals. Ages 21-23 comprise the largest group, and law students account for 250 records (64.9%). Advocates or legal professionals account for 36 (9.4%), while law graduates account for 73 (19.0%). Current occupational status is distinct from the education/professional-level variable; the 53 records labelled professional education must not be substituted for the 36 practising-status records.

Table 2. Complete respondent profile

Characteristic	Recorded categories and counts
Age	21-23: 149; 24-26: 90; 18-20: 67; 31+: 43; 27-30: 36
Gender	Male: 201; Female: 174; Prefer not to say: 6; Other: 4
Current status	Law student: 250; Law graduate: 73; Advocate or legal professional: 36; Other relevant respondent: 26
Education	Undergraduate: 213; Graduate: 80; Professional: 53; Postgraduate: 39
Location	Dhaka: 139; Other city corporation: 90; District town: 83; Upazila-rural area: 64; Other: 9
AI experience	1-2 years: 137; 6-12 months: 99; Less than 6 months: 73; More than 2 years: 65; Never: 11
AI frequency	Weekly: 137; Daily or almost daily: 122; Monthly: 77; Less than monthly: 36; Never: 13
AI training	No: 207; Yes, short orientation: 115; Yes, course or certified training: 63

N = 385. Categories sum to 385 within each field. Current status and education are separate variables. No judges-only or nationally representative sample is inferred.

Dhaka contributes 139 respondents (36.1%); other locations are also represented, but location diversity does not establish national coverage. Any AI training is reported by 178 respondents (46.2%). Daily or near-daily and weekly use together account for 259 records (67.3%). All 20 agreement items are complete and coded within one to five. There are no duplicate identifiers, duplicate complete 20-item response vectors or records giving one uniform answer to all 20

statements. These checks identify observable data properties; they do not establish authenticity, independence or representativeness by themselves.

4.2 Conditional trust and governance expectations

Table 3 presents all five original response categories for the two primary outcomes. Explanation-contingent trust is endorsed by 215 respondents (55.8%), whereas validation-contingent trust is endorsed by 247 (64.2%). The corresponding strong-agreement counts are 46 and 77. Neutral responses remain substantial: 144 for EXT1 and 119 for EXT4. These distributions demonstrate why collapsing every non-endorsing response into distrust would mischaracterise the data.

Table 3. Conditional-trust response distributions

Item	Mean (SD)	Counts 1 / 2 / 3 / 4 / 5	Endorsed n (%)	95% CI %
EXT1	3.61 (0.79)	1 / 25 / 144 / 169 / 46	215 (55.8%)	50.9-60.7
EXT4	3.79 (0.82)	1 / 18 / 119 / 170 / 77	247 (64.2%)	59.2-68.8

N = 385. Categories: one strongly disagree; two disagree; three neutral; four agree; five strongly agree. Endorsement = four or five. Intervals are pointwise Wilson intervals; means summarise numerical coding.

The two outcomes have different descriptive distributions, but neither records a before-and-after trust change. Respondents are agreeing with a statement about what would increase trust, not reporting an experimentally observed increase. Similarly, EXT4's higher mean does not establish that validation is universally more effective than explanation. It establishes a difference between answers to the particular propositions asked in this sample.

Table 4 places those outcomes alongside the four governance expectations and reported verification practice. Review or appeal has the highest endorsement among the four focal expectations: 328 respondents (85.2%). Factor disclosure receives 256 endorsements (66.5%), challengeability 251 (65.2%) and non-reliance on unexplained recommendations 239 (62.1%). Verification practice is endorsed by 339 respondents (88.1%).

Table 4. Governance expectations and verification practice

Item	Mean (SD)	Counts 1 / 2 / 3 / 4 / 5	Endorsed n (%)	95% CI %
EXT2	3.79 (0.76)	0 / 15 / 114 / 193 / 63	256 (66.5%)	61.6-71.0
EXT3	3.76 (0.80)	0 / 23 / 111 / 188 / 63	251 (65.2%)	60.3-69.8
EXT5	3.76 (0.83)	1 / 19 / 126 / 165 / 74	239 (62.1%)	57.1-66.8
ABPF3	4.21 (0.72)	0 / 6 / 51 / 186 / 142	328 (85.2%)	81.3-88.4
AIL2	4.26 (0.67)	0 / 2 / 44 / 190 / 149	339 (88.1%)	84.4-90.9

N = 385. Categories: one strongly disagree; two disagree; three neutral; four agree; five strongly agree. Endorsement = four or five. Intervals are pointwise Wilson intervals; means summarise numerical coding.

The item ordering separates remedial access from other governance expectations. It does not imply that respondents ranked the safeguards in a forced-choice task. Each could be endorsed independently. High agreement with verification also remains a self-report: the database contains no external test of whether respondents detected an incorrect legal citation, checked an underlying source or revised a judgement after verification.

4.3 Paired differences and joint configurations

Table 5 tests the two within-respondent contrasts. For validation versus explanation, 107 respondents endorse only EXT4 and 75 endorse only EXT1. The difference is 32 respondents, or 8.3 percentage points, with a 95% bootstrap interval of 1.6-15.1 points and Holm-adjusted p = .021. For review versus challengeability, the discordant counts are 109 and 32, yielding a 20.0-point difference and adjusted p < .001.

Table 5. Paired endorsement contrasts

Contrast	First only / second only	Difference pp	95% CI pp	Holm p
EXT4 − EXT1	107 / 75	8.3	1.6-15.1	0.021
ABPF3 − EXT3	109 / 32	20.0	14.3-25.7	< .001

N = 385; pp = percentage points. Exact McNemar tests with Holm correction across two contrasts. Pointwise percentile intervals use 10,000 whole-record bootstrap resamples. The statements differ in wording and substantive scope.

The first contrast provides evidence of a sample-level difference between the two stated trust conditions. The second shows stronger endorsement of review or appeal for affected people than of questioning recommendations by users. The wording differs in both institutional mechanism and beneficiary, so that result should not be described as a psychological inconsistency or a clean estimate of preference for one otherwise identical safeguard.

Table 6 reports the full joint configurations. Both trust conditions are endorsed by 140 respondents (36.4%), while 63 (16.4%) endorse neither. For challengeability and review, 219 (56.9%) endorse both and 25 (6.5%) endorse neither. The joint display makes clear that the paired differences coexist with considerable overlap.

Table 6. Joint endorsement configurations

Item pair	Both n (%)	First only n (%)	Second only n (%)	Neither n (%)
Reasons / validation	140 (36.4%)	75 (19.5%)	107 (27.8%)	63 (16.4%)
Challenge / review	219 (56.9%)	32 (8.3%)	109 (28.3%)	25 (6.5%)

N = 385. Rows sum to 385. Endorsement = agree or strongly agree; neither includes neutral responses. The cells are descriptive configurations rather than validated respondent types.

Neither set of cells establishes a latent respondent type. The categories are transparent summaries of two binary indicators, and respondents outside endorsement may be neutral. In particular, the 107 validation-only endorsements should not be labelled rejection of explanations: they include people who chose neutral on EXT1.

4.4 Adjusted ordinal associations

The common-slope diagnostics did not reject proportionality for EXT1 (Wald $\chi^2(16) = 12.17$, $p = .732$) or EXT4 ($\chi^2(16) = 14.58$, $p = .556$). Maximum variance inflation was 1.09, and second-optimiser coefficient differences were below 0.000002. These checks support numerical stability and the chosen proportional-odds summary within the fitted specifications; they do not establish a correct causal model.

Table 7 reports the eight focal coefficients. For EXT1, the adjusted odds ratios range from 0.86 for review or appeal to 1.23 for challengeability. For EXT4, they range from 1.01 for non-reliance on unexplained recommendations to 1.26 for review or appeal. Every pointwise interval includes one, and every Holm-adjusted p-value is at least .817. Thus, H1-H4 lack clear adjusted support for either outcome in this analysis.

Table 7. Focal adjusted ordinal associations

Expectation	EXT1 OR (95% CI)	Holm p	EXT4 OR (95% CI)	Holm p
Factor disclosure	1.11 (0.86-1.43)	1.000	1.12 (0.87-1.45)	1.000
Challengeability	1.23 (0.94-1.60)	0.904	1.02 (0.81-1.29)	1.000
Unexplained non-reliance	1.14 (0.89-1.45)	1.000	1.01 (0.81-1.26)	1.000
Review or appeal	0.86 (0.66-1.12)	1.000	1.26 (0.95-1.67)	0.817

N = 385. OR = proportional odds ratio for a higher outcome response per predictor agreement point. Each model includes all four expectations, verification, training, frequency and age. Categories one and two are pooled in outcomes. Pointwise HC1 intervals; Holm correction across eight focal coefficients.

The imprecision matters substantively. For example, the EXT4 review estimate is compatible with both a small negative association and a positive association of potentially relevant size. The analysis therefore does not establish equivalence, irrelevance of review or absence of all explanation-trust relationships. It shows that these particular governance expectations do not clearly distinguish stronger conditional-trust responses after the specified adjustment and multiplicity control. Figure 2 displays the estimates without converting the hypotheses into causal paths.

Figure 2. Adjusted ordinal associations for two conditional trust propositions. N = 385; pointwise 95% HC1 intervals; eight focal comparisons receive Holm adjustment.

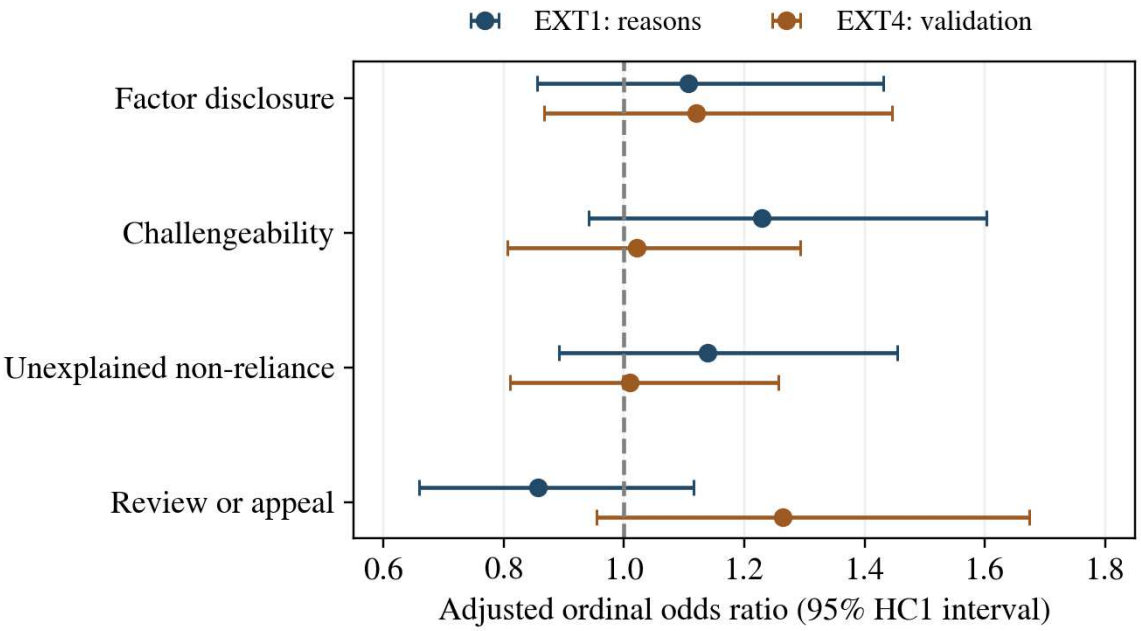


Figure 2. Adjusted ordinal associations for two conditional trust propositions

Table 8 reports contextual coefficients. AI-use frequency aligns positively with both outcomes: ordinal odds ratio 1.28 (95% interval 1.07-1.53) for EXT1 and 1.29 (1.08-1.54) for EXT4. Verification practice has estimates below one for EXT1 and above one for EXT4, but both intervals include one. Training and age also have intervals spanning one. These covariate results are descriptive and are not presented as independently confirmed determinants of trust.

Table 8. Contextual coefficients in the ordinal models

Contextual variable	EXT1 OR (95% CI)	EXT4 OR (95% CI)
Verification practice	0.85 (0.65-1.11)	1.27 (0.93-1.72)
Any AI training	1.35 (0.92-2.00)	1.06 (0.71-1.58)
AI frequency category	1.28 (1.07-1.53)	1.29 (1.08-1.54)
Age 27+	1.24 (0.75-2.07)	0.88 (0.55-1.39)

N = 385. Verification is per agreement point; training compares any with none; frequency is per coded category; age compares 27+ with younger. Pointwise HC1 intervals are descriptive and outside the focal testing family.

The McFadden pseudo-R^2 values are .021 for both models. These are likelihood-based fit summaries, not percentages of variance explained. Taken together, the estimates indicate that the available expectations and contextual covariates account for relatively little improvement in fitted response likelihood over threshold-only models. A more comprehensive explanation

would require better measures of system experience, institutional trust, task risk and actual comprehension.

4.5 Direct outcome comparisons and sensitivity

Table 9 compares the four focal ordinal coefficients across outcomes using their paired covariance. The joint test is $\chi^2(4) = 5.54$, $p = .236$. Review or appeal has a pointwise ratio of odds ratios of 1.47 (95% interval 1.02-2.13), but its four-comparison Holm-adjusted p-value is .151. The nominal interval should not be used to override the adjusted comparison or claim a demonstrated difference between outcomes.

Table 9. Direct comparisons of focal associations across outcomes

Expectation	EXT4 / EXT1 OR ratio (95% CI)	Holm p
Factor disclosure	1.01 (0.70-1.47)	0.950
Challengeability	0.83 (0.59-1.18)	0.898
Non-reliance on unexplained outputs	0.89 (0.64-1.22)	0.909
Review or appeal	1.47 (1.02-2.13)	0.151

N = 385. Ratios exponentiate EXT4 minus EXT1 log-odds coefficients. Covariance retains pairing through differences in per-record influence vectors. Pointwise intervals; Holm correction across four comparisons. Joint Wald $\chi^2(4) = 5.54$, $p = .236$.

Table 10 reports the focal estimates after adding student status and after excluding the 24 inconsistent-use records (n = 361). The pointwise intervals continue to span one for all four expectations in both outcomes. Binary threshold checks likewise yield no focal pointwise interval excluding one. The strong-agreement models have only 46 EXT1 events and 77 EXT4 events; their wider intervals are especially relevant when assessing apparent changes in estimate size.

Table 10. Sensitivity of the focal ordinal estimates

Specification	Expectation	EXT1 OR (95% CI)	EXT4 OR (95% CI)
Student status added	Disclosure	1.11 (0.86-1.43)	1.14 (0.88-1.48)
	Challenge	1.23 (0.94-1.60)	1.02 (0.80-1.29)
	Unexplained non-reliance	1.14 (0.89-1.45)	1.01 (0.81-1.25)
	Review or appeal	0.86 (0.66-1.12)	1.25 (0.94-1.66)
Consistent use	Disclosure	1.05 (0.81-1.37)	1.10 (0.83-1.44)
	Challenge	1.25 (0.93-1.67)	0.94 (0.74-1.21)
	Unexplained non-reliance	1.10 (0.85-1.42)	1.04 (0.83-1.30)
	Review or appeal	0.89 (0.68-1.17)	1.27 (0.95-1.70)

Student-status-added models: N = 385. Consistent-use models: n = 361, excluding all records with never-use frequency/experience inconsistencies. Intervals are pointwise HC1. No new confirmatory hypothesis family is assigned to sensitivity estimates.

The sensitivity results support the narrower conclusion of limited adjusted evidence within these specifications. They do not address unknown recruitment selection, unrecorded clustering or the distinction between hypothetical expectations and behaviour. The most stable contribution remains the paired descriptive separation of conditions and procedures, alongside an explicit account of the uncertainty in their adjusted relationships.

5. Discussion

The findings support a differentiated interpretation of trust conditions in judicial AI. Independent validation receives more endorsement than understanding reasons, and access to review or appeal receives more endorsement than the ability to question a recommendation. Yet these descriptive differences do not translate into clear adjusted relationships between the four governance expectations and the two conditional trust propositions. The discussion explains why these results can coexist, how they relate to prior scholarship and what they imply for an institutional account of appropriate reliance.

5.1 Validation and explanation address different vulnerabilities

The paired validation-explanation difference suggests that respondents distinguish an intelligible recommendation from an independently evaluated tool. This distinction is consistent with trust scholarship's separation of positive expectations from the basis for those expectations. An explanation can make a particular output understandable while leaving unanswered whether the system performs adequately across cases. Validation can address aspects of capability without ensuring that the reasons offered in an individual case are

intelligible. The questionnaire cannot determine whether respondents reasoned in exactly these terms, but its separate items allow the distinction to remain analytically visible.[23]

This interpretation also fits the literature on appropriate reliance. A system should not earn confidence merely by presenting a convincing narrative. Its outputs need to be assessed against task-specific evidence, including the possibility of failure. At the same time, a favourable validation result cannot make every subsequent recommendation correct. The sample's higher endorsement of validation therefore motivates a governance question about how evaluation evidence should accompany explanations; it does not justify replacing explanations with a certificate of reliability.[24]

The data do not establish that validation causes a larger trust increase. EXT1 and EXT4 are not randomised treatments, and their wording invokes different processes. One asks about personal understanding; the other asks about an external assurance activity whose details are unspecified. The observed difference could reflect these features as well as substantive preferences. An experiment would need to vary explanation content and validation evidence independently, specify the evaluator's independence and measure both justified acceptance and justified rejection. Research on explanations and on perceived algorithmic trustworthiness supplies a foundation for such a design.[25]

5.2 Review, challenge and the architecture of procedure

The 20-point difference between review or appeal and challengeability is substantively striking because it concerns two related procedural expectations. However, the distinction should be interpreted through their wording. ABPF3 names people affected by a decision and an institutional remedy. EXT3 names users and the ability to question a recommendation. A respondent could support an appeal while remaining uncertain about what questioning an AI recommendation would involve. This reading is compatible with procedural justice and reviewability scholarship, but the survey does not measure the reasoning behind the response.[26]

The governance implication is to connect opportunities for inquiry with meaningful routes for correction. Disclosure alone gives no assurance that an affected person can obtain relevant records, challenge an erroneous input or receive an effective response. Bovens' account of accountability helps identify the missing institutional relationships, while Ananny and Crawford caution against equating visibility with usable understanding. The current findings add descriptive evidence that these functions should remain separate in empirical instruments and implementation planning. They do not prove that any particular remedy already exists or works in Bangladesh.[27]

---

[23] Mayer, Davis and Schoorman (n 8); Rousseau and others (n 8); Jacovi and others (n 10).

[24] Lee and See (n 3); Hoff and Bashir (n 3); Raji and others (n 12).

[25] Gregor and Benbasat (n 6); Miller (n 6); Donghee Shin, 'The Effects of Explainability and Causability on Perception, Trust, and Acceptance: Implications for Explainable AI' (2021) 146 International Journal of Human-Computer Studies 102551; Stephan Grimmelikhuijsen, 'Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making' (2023) 83 Public Administration Review 241.

[26] Tyler (n 13); Cobbe, Lee and Singh (n 14).

[27] Bovens (n 11); Ananny and Crawford (n 11).

Nor should high support for appeal be interpreted as permission to deploy a system provided an appeal route is advertised. A remedy may be inaccessible, delayed or unable to examine the model contribution. Human oversight can also become formal rather than substantive when reviewers lack time, information or authority. The findings therefore support specifying review conditions, not relying on a generic assurance that a human remains responsible. The relevant institutional question is who can inspect what information and what action follows when a problem is established.[28]

5.3 Descriptive salience does not establish predictive alignment

The weak adjusted findings qualify any claim that explanation-related expectations straightforwardly predict trust. A safeguard can be broadly valued without explaining differences in a conditional trust response. Both outcomes also contain substantial neutrality, and the available predictors cover only a narrow set of reasons respondents might distinguish them. Model fit and uncertainty consequently favour a restrained conclusion: this instrument identifies several salient expectations but provides limited evidence about the adjusted response relationships proposed in H1-H4. That conclusion is narrower than either validating or rejecting explainable AI as a general theory.[29]

Measurement is central to the interpretation. EXT1 already places understanding and trust within the same statement, so a regression of a total explainability score on a total trust score drawn from the same block would risk circularity. By retaining the item as a conditional proposition, the analysis avoids claiming to estimate an independent explanation-to-trust effect. Scale-development and formative-measurement literature support treating content and measurement assumptions as substantive choices rather than selecting them to obtain a preferred model.[30]

Multiplicity also changes the interpretation of apparent differences. The paired-outcome review coefficient has a pointwise interval above one when expressed as an odds-ratio ratio, but its adjusted p-value is .151 and the global test is inconclusive. Highlighting that single nominal result would create a more confident narrative than the family of comparisons supports. Equally, large adjusted p-values do not establish that all associations are negligible. The intervals should guide subsequent precision and design decisions, consistent with the ASA statement and the distinction between missing evidence and evidence of absence.[31]

5.4 Bangladesh, judicial governance and development

For Bangladesh, the results are most useful as evidence about an early legal-education and stakeholder setting. The predominantly student composition directs attention to what future practitioners need to learn: distinguishing understandable language from valid reasoning, checking legal sources, and recognising the difference between evaluation evidence and an unconditional assurance. The positive frequency coefficients do not demonstrate that

[28] Green (n 15); Alon-Barkat and Busuioc (n 15); Busuioc (n 12).

[29] Glikson and Woolley (n 9); Jacovi and others (n 10).

[30] Boateng and others (n 17); Diamantopoulos and Winklhofer (n 17).

[31] Holm (n 19); Wasserstein and Lazar (n 22); Altman and Bland (n 22).

increasing use would improve trust or competence. More frequent users may differ in unmeasured ways, and the outcome remains agreement with a prospective statement.[32]

The international frameworks offer a vocabulary for turning these distinctions into institutional questions. UNESCO's judicial guidance and the CEPEJ charter place transparency alongside rights, oversight and user control. Their comparative value lies in this plurality of functions. The present article applies that insight to the survey rather than importing either instrument as binding Bangladeshi law. The national draft-policy context similarly permits discussion of options without claiming that consultation text establishes an enforceable rule.[33]

These questions connect to SDG 16 through access to justice and accountable institutions. The survey does not measure progress towards an SDG target. Its relevance is more specific: AI governance should be assessed through the ability to examine and contest decisions, rather than through an adoption count or a favourable attitude score alone. A system that attracts confidence while reducing effective scrutiny would not meet that institutional objective. This is the article's normative inference from the combined framework, not an outcome directly measured in the database.[34]

5.5 Limits and directions for further research

The study is limited by undocumented recruitment details, unknown clustering and a sample that cannot represent judicial actors nationally. Single-item outcomes do not separate baseline trust, comprehension, warranted confidence and reliance behaviour. Common-source responses may reflect response styles or shared wording rather than substantive alignment. The audit and robust covariance calculations do not remove these limitations. No causal explanation should be built from the small pseudo-R^2 values or from one contextual coefficient.[35]

A subsequent study should retain the present distinctions while improving design. It could recruit judges, advocates, court staff and affected parties through documented strata; measure trust before exposure; vary explanation and validation evidence independently; and include correct and incorrect recommendations. Outcomes should include source checking, error detection, reliance and willingness to challenge. Separate measures of procedural fairness and review accessibility would permit stronger theoretical tests. The present findings justify this programme by identifying distinct questions and uncertain relationships, rather than by claiming a completed validation of the explanation-trust mechanism.

6. Theoretical and Practical Implications

The implications follow from the combination of descriptive separation and uncertain adjusted relationships. A useful judicial AI framework must distinguish why a recommendation is intelligible, what evidence supports its reliability and how an affected person can obtain

[32] Davis (n 16); Hoff and Bashir (n 3).

[33] Gutiérrez (n 2); CEPEJ (n 2); ICT Division, Draft AI Policy (n 4).

[34] United Nations General Assembly, *Transforming Our World: The 2030 Agenda for Sustainable Development* (25 September 2015) UN Doc A/RES/70/1, targets 16.3 and 16.6; Kroll and others (n 7).

[35] Philip M Podsakoff and others, 'Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies' (2003) 88 Journal of Applied Psychology 879.

correction. The study contributes to that distinction while leaving causal mechanisms open. Its practical recommendations are prospective design considerations for institutions and researchers; they are not assertions that respondents have evaluated a deployed system or that the database establishes the effectiveness of a particular governance intervention.

6.1 A layered account of warranted reliance

The theoretical contribution is a layered account of conditional trust. The first layer concerns comprehensible reasons, the second independent evidence about system performance, and the third procedural opportunities to question and obtain review. These layers can reinforce one another without forming a single psychological construct. An accessible explanation may reveal a limitation requiring external evaluation; a validation result may identify settings requiring additional scrutiny; a review procedure may require records that make the system's contribution examinable. This account connects explanation research to institutional accountability rather than treating favourable confidence as the final outcome.[36]

The results also clarify the relation between descriptive and explanatory contributions. Strong endorsement of review does not imply that review expectations explain variation in conditional trust. Equally, an uncertain regression coefficient does not reduce the normative importance of review. Keeping these inferential levels separate permits empirical work to inform governance without giving a survey the authority to settle a rights question. That distinction is especially relevant where respondents are predominantly students and the institutional consequences would fall on people outside the sample.

A further implication concerns measurement. Future instruments should separate exposure to an explanation, comprehension of that explanation, perceived explanation quality, baseline trust and observed reliance. Validation should specify who conducts it, what evidence is available and which tasks and populations it covers. Challengeability and appeal should identify the person entitled to act and the institution responsible for responding. These distinctions would permit a genuine test of explanation-to-trust pathways while avoiding the semantic overlap present in the current block.[37]

6.2 Institutional design and professional education

For prospective judicial procurement, an explanation requirement should identify intended audiences and decision tasks. An administrator's performance summary, a practitioner's source trace and an affected party's account of how an output entered a decision serve different purposes. Procurement documents can specify these purposes and require evidence that the relevant audience can use the information. The current survey supplies a reason to investigate such arrangements; it does not determine which explanation format will work best.[38]

Independent validation should similarly be tied to an explicit scope. A provider's general accuracy claim is not equivalent to evaluation on the legal materials, languages and tasks for which a court proposes to use the system. Evaluation should consider errors, uncertainty and

[36] Miller (n 6); Bovens (n 11); Cobbe, Lee and Singh (n 14).

[37] Boateng and others (n 17); Jacovi and others (n 10).

[38] Gregor and Benbasat (n 6); Ananny and Crawford (n 11).

conditions under which use should stop or require additional checking. Governance arrangements should specify responsibility for reassessment after material changes. These recommendations follow from the distinction between a persuasive output and a dependable process, supported by auditing and accountability scholarship.[39]

Review arrangements should preserve sufficient information to assess the AI contribution, including the material considered, relevant system version and the human decision-maker's treatment of the output, subject to lawful confidentiality protections. Affected people need an intelligible route for raising an objection and an accountable body capable of responding. The article does not prescribe a new domestic legal right from the survey; it identifies questions institutions should resolve through applicable law and legitimate rule-making processes.[40]

Legal education can teach these functions through exercises that require students to evaluate both sound and flawed AI recommendations. Assessment should reward appropriate rejection as well as justified acceptance. Students can be asked to distinguish an explanation from evidence, identify what additional validation is needed and describe how an error could be challenged. Such exercises directly address the sample's educational composition while avoiding the unsupported inference that more frequent AI use necessarily produces competence. Their effectiveness should be tested rather than assumed.[41]

7. Conclusion

This study examined explanation and independent validation as conditions for trust in judicial AI using 385 adult, AI-aware survey records from Bangladesh. It retained the distinction between respondents' expectations and the behaviour of an actual system. By analysing items separately, it avoided treating mixed normative and conditional wording as a validated explainability-trust scale. The result is a completed empirical account of what the questionnaire can establish, together with a clear boundary around the stronger explanation-to-trust mechanism that it cannot identify.

The central descriptive finding is that independent-validation-contingent trust receives greater endorsement than explanation-contingent trust, by 8.3 percentage points in a paired comparison. Access to review or appeal also receives greater endorsement than the ability to question recommendations, by 20.0 points. The joint distributions show considerable overlap rather than opposing respondent camps. These patterns support treating reasons, evaluation and remedies as distinct questions in both research instruments and governance planning. They do not demonstrate that one safeguard is causally superior to another.

The adjusted models provide a complementary qualification. None of the eight focal governance-expectation associations receives clear support after Holm adjustment, and the joint paired-outcome comparison does not establish different association patterns across the two trust conditions. Sensitivity analyses do not materially change that conclusion. The findings therefore offer limited evidence for the proposed response relationships, rather than

[39] Raji and others (n 12); Busuioc (n 12); Kroll and others (n 7).

[40] Cobbe, Lee and Singh (n 14); CEPEJ (n 2).

[41] Lee and See (n 3); Green (n 15).

proving that explanations or review are unimportant. Their uncertainty should inform subsequent study design and discourage the selective promotion of isolated nominal results.

The article's theoretical contribution is a layered account of warranted reliance: intelligible reasons, independently examined capability and effective opportunities for challenge and correction address different vulnerabilities. Its practical contribution is to specify institutional questions arising from those layers. Who needs an explanation, what evidence validates the tool for a particular task, what information is retained, and who can act when a problem is identified? These questions connect AI governance to accountable decision-making without using favourable attitudes as a substitute for institutional evidence.

The study's scope remains a predominantly student sample with undocumented recruitment details and single-item conditional outcomes. Further research should involve distinct judicial and public constituencies, documented recruitment and behavioural tasks. Experiments should vary explanations and validation evidence independently while testing the recognition of errors and the appropriate rejection of unreliable advice. Such a programme would move from stated expectations to evidence about calibrated reliance. The present study provides a transparent starting point by identifying meaningful distinctions, quantifying their endorsement and preserving uncertainty about their relationships.

Declarations

Ethical approval and oversight

Ethical approval for this study was obtained from the Center for Academic & Professional Career Development and Research (CAPCDR), an external institutional body, under approval reference CAPCDR: SMUCT/LLB/SS-20261012-01. External ethical review was sought because Shanto-Mariam University of Creative Technology does not have an institutional ethics committee responsible for reviewing this type of research.

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

Maksuda Ferdous: 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

Alon-Barkat S and Busuioc M, 'Human-AI Interactions in Public Sector Decision Making: "Automation Bias" and "Selective Adherence" to Algorithmic Advice' (2023) 33 Journal of Public Administration Research and Theory 153

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

Ananny M and Crawford K, 'Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability' (2018) 20 New Media & Society 973

Arrieta AB and others, 'Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI' (2020) 58 Information Fusion 82

Boateng GO and others, 'Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer' (2018) 6 Frontiers in Public Health 149

Bovens M, 'Analysing and Assessing Accountability: A Conceptual Framework' (2007) 13 European Law Journal 447

Brant R, 'Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression' (1990) 46 Biometrics 1171

Busuioc M, 'Accountable Artificial Intelligence: Holding Algorithms to Account' (2021) 81 Public Administration Review 825

Cobbe J, Lee MSA and Singh J, 'Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems' in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2021) 598

Dahl M and others, 'Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models' (2024) 16 Journal of Legal Analysis 64

Davis FD, 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology' (1989) 13 MIS Quarterly 319

Diamantopoulos A and Winklhofer HM, 'Index Construction with Formative Indicators: An Alternative to Scale Development' (2001) 38 Journal of Marketing Research 269

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

European Commission for the Efficiency of Justice, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment* (Council of Europe 2018)

Glikson E and Woolley AW, 'Human Trust in Artificial Intelligence: Review of Empirical Research' (2020) 14 Academy of Management Annals 627

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

Gregor S and Benbasat I, 'Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice' (1999) 23 MIS Quarterly 497

Grimmelikhuijsen S, 'Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making' (2023) 83 Public Administration Review 241

Gutiérrez JD, *Guidelines for the Use of AI Systems in Courts and Tribunals* (UNESCO 2025)

Hoff KA and Bashir M, ‘Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust’ (2015) 57 Human Factors 407
Holm S, ‘A Simple Sequentially Rejective Multiple Test Procedure’ (1979) 6 Scandinavian Journal of Statistics 65
Information and Communication Technology Division, *National Artificial Intelligence Policy 2026-2030* (Draft Version 2.0, 2026) <https://aipolicy.gov.bd/docs/national-ai-policy-bangladesh-2026-2030-draft-v2.0.pdf> accessed 26 October 2026
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
Kroll JA and others, ‘Accountable Algorithms’ (2017) 165 University of Pennsylvania Law Review 633
Lee JD and See KA, ‘Trust in Automation: Designing for Appropriate Reliance’ (2004) 46 Human Factors 50
Mayer RC, Davis JH and Schoorman FD, ‘An Integrative Model of Organizational Trust’ (1995) 20 Academy of Management Review 709
McCullagh P, ‘Regression Models for Ordinal Data’ (1980) 42 Journal of the Royal Statistical Society Series B 109
McNemar Q, ‘Note on the Sampling Error of the Difference between Correlated Proportions or Percentages’ (1947) 12 Psychometrika 153
Miller T, ‘Explanation in Artificial Intelligence: Insights from the Social Sciences’ (2019) 267 Artificial Intelligence 1
Podsakoff PM and others, ‘Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies’ (2003) 88 Journal of Applied Psychology 879
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
Rousseau DM and others, ‘Not So Different After All: A Cross-Discipline View of Trust’ (1998) 23 Academy of Management Review 393
Shin D, ‘The Effects of Explainability and Causability on Perception, Trust, and Acceptance: Implications for Explainable AI’ (2021) 146 International Journal of Human-Computer Studies 102551
Stefanski LA and Boos DD, ‘The Calculus of M-Estimation’ (2002) 56 American Statistician 29
Surden H, ‘Artificial Intelligence and Law: An Overview’ (2019) 35 Georgia State University Law Review 1305
Tyler TR, *Why People Obey the Law* (Yale University Press 1990)
United Nations General Assembly, *Transforming Our World: The 2030 Agenda for Sustainable Development* (25 September 2015) UN Doc A/RES/70/1, targets 16.3 and 16.6
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. Analytical item wording

The following table reproduces the exact wording used to interpret the seven analytical agreement items. It provides an audit trail for the distinction between conditional trust, normative expectations and reported verification.

Table A1. Exact analytical item wording

Code	Statement
EXT1	I trust an AI recommendation more when I can understand the main reasons behind it.
EXT2	AI-assisted legal decisions should disclose the important factors used to reach a recommendation.
EXT3	Users should be able to question or challenge AI-supported recommendations.
EXT4	Independent validation would increase my trust in AI decision-support tools.
EXT5	Unexplained AI recommendations should not be relied upon in high-stakes legal decisions.
ABPF3	People affected by AI-supported decisions should have access to review or appeal.
AIL2	I verify important AI-generated information before relying on it.

Source: Group 3 Codebook. Responses range from one (strongly disagree) to five (strongly agree). These items are not combined into a total score.

The remaining profile fields retain the categories reported in Table 2. No item has been rewritten for analysis or reverse coded.