JOURNAL OF POLICIES AND RECOMMENDATIONS

Volume: 5 Issue: 4

(October-December), 2026

ISSN: 2957-4129



KMF PUBLISHERS

www.kmf-publishers.com/jopar/

Contents lists available at KMF Publishers

KMF Journal of Policies and Recommendations

journal homepage: www.kmf-publishers.com/jopar/

Research Article

Gender Bias Concerns and Support for Algorithmic Accountability in Bangladesh

Surove Syed Oaishe[1]; Md. Didarul Islam[1]; Maruf Hasan Joy[1]; Kazi Abdul Mannan[2]

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

Corresponding Author Surove Syed Oaishe, Email: suroveoaishe69@gmail.com

ARTICLE INFO

Keywords:
gender bias; algorithmic accountability; audit; procedural justice; remedy; Bangladesh.

ABSTRACT

Governance of gender bias in artificial intelligence (AI) involves distinct commitments: testing for unequal outcomes, assigning responsibility, explaining safeguards and providing remedies. An overall measure of support can obscure differences among these commitments. This cross-sectional study analysed 385 adult responses to a common questionnaire on AI, gender, equality and law in Bangladesh. We examined individual response distributions, compared endorsements within respondents and estimated exploratory adjusted associations with two separate outcomes: support for effective remedy and anticipated confidence from transparency about bias testing. Effective remedy and organisational responsibility were each endorsed by 302 respondents (78.4%); testing for gender-related discriminatory outcomes was endorsed by 215 (55.8%). The paired remedy-testing difference was 22.6 percentage points (95% CI 16.1 to 29.1; Holm-adjusted $p < 0.001$). Both were endorsed by 167 respondents (43.4%). Bias concern was associated with anticipated confidence from transparency ($b = 0.204$, 95% CI 0.096 to 0.312; adjusted $p < 0.001$), but the corresponding association with remedy endorsement on the original rating scale was imprecise and near zero. These findings distinguish endorsement of redress from endorsement of a preventive organisational practice. They also distinguish anticipated reassurance from observed changes in trust. The sample is concentrated in legal study and practice and does not support population prevalence estimates. The findings motivate separate evaluation of testing, communication and access to remedy.

Received: 20 September 2026 / Revised: 25 October 2026 / Accepted: 28 October 2026 / Published: 30 October 2026

Copyright: © 2026 by the authors. Licensee KMF Publishers (www.kmf-publishers.com). This open-access article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
DOI: https://doi.org/10.64907/xkmf.v5i4.jopr.4

1. Introduction

Automated tools can assist people in finding information and help organisations rank, classify or review cases. The same tools can influence which people receive attention, which applications are selected for review and what evidence is available to challenge a result. Their effects depend on the setting: a language model that helps a user understand a form raises different concerns from a scoring system used to rank applicants. In either case, an error may carry greater costs for people who have fewer alternatives or less access to an explanation. Assessing fairness requires specifying the decision, the people affected and the institution that can correct it.[1]

Gender-related bias may arise through the choice of training data, labels, evaluation targets or operational procedures. A model need not explicitly use gender to produce uneven outcomes. Other information can correlate with gendered patterns, and an apparently uniform procedure may work differently for people whose circumstances are poorly represented. A well-known audit of commercial gender classification found substantial differences in performance across groups defined by skin type and gender. Its results illustrate the value of disaggregated evaluation in that application. They do not establish that all systems show the same disparities or that the categories and procedures used there are suitable for every deployment.[2]

An organisation considering a system can test for unequal outcomes before deciding to use it. Once the system is deployed, monitoring can detect changes that an initial evaluation missed. After a disputed decision, an affected person may need an explanation, a reconsideration or another form of remedy. These tasks support a shared goal but require different evidence and authority. An audit may identify a pattern across many cases without resolving an individual complaint. A complaint can reveal a failure in one case without showing whether it recurs across a population. A useful accountability design therefore connects the stages without treating any one of them as a complete answer.[3]

An opinion survey can face an analogous problem. Asking whether someone supports algorithmic accountability does not identify whether they favour testing, a duty to act on findings, transparency about procedures, or a route to remedy. If a study averages these statements together, a large difference in endorsement may be hidden. Likewise, a person can endorse an effective remedy regardless of whether they feel especially concerned about AI bias. The widespread endorsement of a remedy is a separate observation from an association between remedy ratings and concern.

Articles 27 and 28 of the Constitution of Bangladesh provide a domestic equality

[1] Andrew D Selbst and others, ‘Fairness and Abstraction in Sociotechnical Systems’ (Proceedings of the Conference on Fairness, Accountability, and Transparency, ACM 2019) 59.

[2] Joy Buolamwini and Timnit Gebru, ‘Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification’ (2018) 81 Proceedings of Machine Learning Research 77.

[3] Inioluwa Deborah Raji and others, ‘Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing’ (Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, ACM 2020) 33; Mark Bovens, ‘Analysing and Assessing Accountability: A Conceptual Framework’ (2007) 13 European Law Journal 447.

context. Article 27 concerns equality before law and equal protection; article 28 addresses specified forms of discrimination and equal rights of women in spheres of State and public life. These provisions frame important questions about the use of automated tools in public decision-making. This study does not infer a universal private-sector audit obligation or a new cause of action from them. Which requirements apply to a particular deployment depends on its legal and factual setting. Survey preferences can inform discussions about institutional procedures, but cannot establish their legal status.[4]

The Group 10 questionnaire also includes questions about AI-enabled legal-service adoption and women's legal empowerment. The present analysis uses its gender-bias and accountability block. Its focus is the relationship between concern about reproduction of bias and support for specific safeguards, rather than comparison of adoption by gender or assessment of empowerment. The shared data source should be disclosed when findings from different Group 10 papers are considered together.

First, how often do respondents endorse bias concern, organisational testing, effective remedy, organisational responsibility and anticipated confidence from transparency? Second, how often do they endorse both testing and remedy, and how large are the paired endorsement differences when testing is used as a benchmark? Third, are bias concern and endorsement of organisational duties associated with remedy ratings or anticipated confidence after adjustment for recorded background differences?

The article's contribution is to separate preventive scrutiny from remedial expectations and to show which positions coexist within individuals. This approach provides a more direct basis for selecting consultation questions than a single accountability score. The secondary analysis separates expected reassurance from support for relief, so an association with one is not interpreted as evidence about the other. The analysis concerns attitudes in an AI-aware adult sample, not the fairness of a named product or the efficacy of an institutional intervention.

2. Literature Review and Theoretical Foundation

This review links gender and technology scholarship with procedural justice and accountability. The first examines how technical design and social arrangements interact. The second asks whether people have meaningful opportunities to understand and question a decision. Accountability identifies which actors are expected to explain and respond. The questionnaire does not measure complete theories; the framework is used to clarify the roles of the observed items and the limits of claims based on them.

[4] Constitution of the People’s Republic of Bangladesh, arts 27–28.

2.1 Gender bias as a problem of systems and institutions

Gender bias can refer to several distinct phenomena: systematically inaccurate classification, unequal access to a service, or a decision process that reproduces a disadvantage already present in its setting. These phenomena can overlap, but they do not always have the same cause or solution. For example, improving the accuracy of classification does not determine whether the classification is necessary. Equal error rates do not by themselves establish that the objective of a ranking process is appropriate. A governance assessment should therefore begin by specifying why the tool is used, how outputs are acted on and who bears the costs of error.[5]

The gender and technology perspective also directs attention to what institutions choose to observe. A provider may measure performance for categories it has recorded but miss people who are poorly represented, whose identities are misclassified or who cannot access the service. Disaggregation can expose a problem if the data and categories are appropriate, but can also create privacy risks and unstable estimates when groups are small. A testing plan needs a defensible purpose for collecting information, evidence that a category is meaningful in context, and a way to report results without exposing individuals. It should make clear which outcomes were measured and which experiences remain outside the evaluation.[6]

GBA1 asks whether respondents are concerned that AI systems can reproduce gender bias found in data or institutions. Agreement records a judgement about a potential mechanism. It does not demonstrate that a respondent has experienced harm, can detect a disparity in a model or believes that all systems discriminate. It also does not identify which technical or organisational intervention the respondent thinks appropriate. The item is therefore interpreted as bias concern rather than measured exposure, risk probability or objective knowledge.

2.2 Preventive testing and organisational responsibility

Testing is a structured attempt to identify failures before they become routine. It can include evaluating performance across relevant groups, examining how decisions change when system outputs are used and checking whether remedies are available for foreseeable errors. Testing during development can reveal defects, while testing in the deployment context can show whether assumptions hold for the intended population and workflow. An internal auditing framework can preserve evidence across the life cycle of a system and identify who must act on a finding. Its value depends on whether results can change procurement, deployment or operation.[7]

GBA2 refers specifically to testing gender-related discriminatory outcomes. It does not describe the methods, disclosure requirements, costs or timing of a test. A

[5] Selbst and others (n 1); Buolamwini and Gebru (n 2).

[6] Buolamwini and Gebru (n 2); National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1, 2023) doi:10.6028/NIST.AI.100-1.

[7] Raji and others (n 3); NIST, AI RMF (n 6).

respondent might support the principle and remain uncertain about the wording because it leaves the institution's role or the form of testing unspecified. Conversely, a positive answer does not tell the researcher whether the respondent favours external audit, internal assessment or publication of a specific fairness metric. The empirical item should therefore be interpreted as support for the stated responsibility to test, with those details open.

GBA4 asks whether developers and deploying organisations should be accountable for preventable discriminatory harms. Developers may control design and documentation, while deployers may determine the setting, the human review process and the consequences of outputs. A workable institutional process can assign duties to both without assuming identical responsibility. In any actual claim, liability would require separate legal analysis. The study keeps the general responsibility statement separate from testing in descriptive reporting. Their mean in regression is an observed summary of organisational duties rather than an estimate of a validated latent trait.[8]

2.3 Procedural justice and effective remedy

A procedure offers voice when affected people can describe what happened and have their account considered. In automated decision settings, access to voice may depend on whether the person knows that a system materially contributed, has a way to obtain relevant information and can reach someone authorised to act. Neutral review and an intelligible explanation can make a challenge meaningful. An institution can acknowledge a concern while still offering no remedy if the reviewer cannot change an erroneous result or refer the person to someone who can. A remedy process therefore needs to be evaluated by what it can do, not by the existence of a contact form.[9]

GBA3 refers to an effective remedy for people affected by biased AI decisions. Its wording does not identify the form of relief or the forum that should provide it. Reconsideration, correction of records, restoration of access and compensation can address different harms. A person's support for remedy is compatible with disagreement about which of these responses is appropriate. The item is a direct indicator of normative endorsement and is not used as evidence of experience with a complaint channel. It also does not measure how people would weigh speed, cost and independence when choosing among possible processes.

A digital inclusion perspective explains why the formal availability of review may not be

[8] Bovens (n 3); Raji and others (n 3).

[9] Bovens (n 3); Mary J Culnan and Pamela K Armstrong, ‘Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation’ (1999) 10 Organization Science 104.

enough. A route can be difficult to use if it requires uninterrupted internet access, specialised language or documents the person does not possess. A useful inquiry should ask what would be required to use the route, not just whether it exists. The questionnaire's empowerment items concern related themes, but they are not part of the focal empirical analysis. Their existence in the workbook does not justify claiming that this analysis measures access barriers or legal empowerment.

2.4 Transparency and conditional confidence

Transparency about testing can explain what a provider examined, which groups or contexts were included and how problems were handled. Information can be useful even when it reduces confidence in a particular tool by revealing uncertainty or an unresolved problem. Appropriate reliance therefore requires attention to the substance of a disclosure, not simply to whether it increases reassurance. A summary that announces a system has passed a fairness test without identifying the outcome, benchmark or limits may be difficult to evaluate. Conversely, a detailed report may be useful to a technical reviewer yet inaccessible to a person affected by a decision.[10]

GBA5 asks whether transparency about bias testing would increase the respondent's confidence. It does not give respondents test results, name a provider or measure confidence before and after a disclosure. A positive response is best understood as anticipated reassurance under the stated condition. It is possible for people who are more concerned to see greater value in disclosure while remaining less confident in existing systems. It is also possible for those who support a remedy to regard access to it as an entitlement independent of confidence. These possibilities motivate separate analyses of the two outcomes.

2.5 Conceptual framework and exploratory expectations

Figure 1 identifies the two focal explanatory measures and the distinct outcomes. Bias concern may track demand for an effective remedy because people who consider harm plausible may place greater weight on relief. Support for remedy may also be widespread for reasons of fairness that are not closely related to personal concern. Bias concern could track anticipated confidence from transparency if respondents consider evidence about testing valuable. Common wording about bias could contribute to that association. Because no condition was manipulated and the dataset was already collected, these propositions guide exploratory analysis rather than confirmatory causal testing.

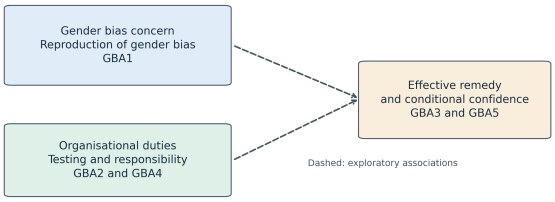

Figure 1. Analytical framework for bias concern and accountability judgements

The organisational duties mean combines support for testing and responsibility for preventable harm. It can vary independently

[10] John D Lee and Katrina A See, 'Trust in Automation: Designing for Appropriate Reliance' (2004) 46 Human Factors 50; Raji and others (n 3).

of bias concern. The models include a general AI orientation summary for context. That summary is based on recognition of inaccuracy and bias and reported verification. It is not a measure of demonstrated competence, and the general oversight item is omitted because its policy wording overlaps with the focal accountability questions. The primary descriptive comparison remains the difference between support for a remedy and support for testing within the same people.

3. Methodology

The study used a cross-sectional questionnaire and analysed the original item ratings. The reporting distinguishes the observed sample from a target population and the recorded consent indicators from formal ethical review. The main analyses preserve the relation between responses made by the same participant.[11]

3.1 Source and analytical sample

The Group 10 workbook contains a Data sheet, a README linking the questionnaire to three research titles and a codebook. The study analysed 385 records. Each included affirmative consent, adult eligibility and general AI awareness. The available records do not establish a probability sampling frame, a verified invitation denominator, exact fieldwork dates or mode of questionnaire distribution. Consequently, no response rate, design weight or nationally representative margin of error is reported. The achieved sample size is an analytical count rather than evidence of random recruitment.

3.2 Screening

We checked record sequence uniqueness, completeness of analysed fields, screening responses and permissible values for all 20 agreement items. No numerical or profile fields were blank, and no duplicate response and profile patterns were found after excluding sequence and contextual text. All agreement ratings were in the 1-5 range. “None” was retained as a valid tool response. The authors confirmed that the numerical responses were empirical and identified the synthetic descriptions attached to the sequence and contextual fields as metadata errors. Contextual text was excluded from analysis and is not quoted.

[11] Erik von Elm and others, ‘The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies’ (2007) 4 PLOS Medicine e296.

3.3 Measurement and analytical roles

Table 1 distinguishes concern, preventive testing, remedy, responsibility and anticipated confidence. Ratings retain their original direction. GBA2 and GBA4 enter the secondary regression as an equal-weight organisational duties mean, while both items are reported separately in the descriptive results. GBA3 and GBA5 are separate single-item outcomes and do not enter the explanatory mean. The adjusted models include the mean of AIL1-3 as a summary of general AI orientation.

Table 1. Analytical roles of accountability and adjustment items

Role	Code	Interpretation
Bias concern	GBA1	Concern about reproduction of gender bias
Bias testing	GBA2	Support for testing gender-related discriminatory outcomes
Remedy	GBA3	Support for effective remedy after biased decisions
Organisational responsibility	GBA4	Support for responsibility for preventable discriminatory harms
Conditional confidence	GBA5	Expected confidence increase from transparency about bias testing
Adjustment	AIL1-3 and profile	Observed AI orientation; status, location and training

GBA2 and GBA4 form the observed organisational duties mean. Outcomes do not enter either predictor.

Appendix A reproduces the wording and anchors. The GBA block is not scored as a common accountability scale. A coefficient alpha and pairwise rank correlations are reported as descriptive diagnostics. Alpha cannot determine dimensionality and does not establish a formative model. The choice to retain separate items follows the distinction between concern, expectations about procedures and a response about confidence. No latent structural or mediation model is estimated.[12]

3.4 Endorsement and paired analyses

We report all five response categories and define endorsement as agreement or strong agreement. Wilson intervals describe individual proportions under a respondent-level working model. The joint table uses GBA2 and GBA3 and reports percentages of the full sample. Non-endorsement includes neutrality as well as disagreement; it is not treated as opposition. The paired comparisons use testing as the common benchmark for the other four GBA statements. Each difference is the number endorsing only the comparator minus the number endorsing only testing, divided by 385.[13]

Exact McNemar tests condition on the discordant responses. Pointwise percentile intervals were obtained from 10,000 resamples of whole respondents using seed 20261103. Whole-record resampling preserves within-person dependence. Holm adjustment covers the four paired tests. The

[12] Klaas Sijtsma, ‘On the Use, the Misuse, and the Very Limited Usefulness of Cronbach’s Alpha’ (2009) 74 Psychometrika 107.

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

comparison concerns responses to statements that differ in wording and purpose. It does not identify choices under a fixed budget or ask respondents to select only one safeguard.[14]

3.5 Exploratory adjusted models

The linear models use original ratings for remedy and conditional confidence, respectively. Both include bias concern, the organisational duties mean, the observed AI orientation mean, and categorical terms for status, location and training. Coefficients describe average differences on a 1-5 scale under the fitted specification. HC3 standard errors address heteroskedasticity. The four focal coefficients across the two outcomes constitute one Holm family; reported intervals are pointwise. These covariates describe observed differences and are not assumed to form a sufficient causal adjustment set.[15]

Sensitivity analyses focus on remedy. We fit a logistic model of remedy endorsement with sandwich standard errors and repeat the linear model among law students. The constant status category contributes no contrasts in that subset. Each sensitivity model has a separate two-term Holm family. The logistic outcome asks about the endorsement threshold rather than the original rating. The student analysis shares records with the main sample and is not a replication or an interaction test.[16]

3.6 Ethics and data protection

All analysed records affirm consent and adult eligibility. This was an adult non-interventional questionnaire; the analysis excluded contextual text and reported aggregate results.

4. Results

The results report sample composition and the distribution of each item, then show the joint testing and remedy pattern and adjusted associations. All primary analyses use the same 385 records.

4.1 Respondent profile

Table 2 shows age, status, location and training. The largest age group was 21-23 (177, 46.0%). Of the respondents, 243 were law students, 54 law graduates, 53 advocates or legal professionals and 35 other relevant respondents. Dhaka accounted for 149 records, district towns for 87 and other city corporations for 77. No formal AI training was reported by 218 respondents.

[14] Quinn McNemar, ‘Note on the Sampling Error of the Difference between Correlated Proportions or Percentages’ (1947) 12 Psychometrika 153; Bradley Efron and Robert J Tibshirani, An Introduction to the Bootstrap (Chapman & Hall 1993); Sture Holm, ‘A Simple Sequentially Rejective Multiple Test Procedure’ (1979) 6 Scandinavian Journal of Statistics 65.

[15] Halbert White, ‘A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity’ (1980) 48 Econometrica 817; Holm (n 14).

[16] David W Hosmer Jr, Stanley Lemeshow and Rodney X Sturdivant, Applied Logistic Regression (3rd edn, John Wiley & Sons 2013); Ronald L Wasserstein and Nicole A Lazar, ‘The ASA Statement on p-Values: Context, Process, and Purpose’ (2016) 70 American Statistician 129.

Table 2. Sample profile

Characteristic	Category	n (%)
Age	21-23	177 (46.0%)
Age	24-26	83 (21.6%)
Age	18-20	64 (16.6%)
Age	27-30	31 (8.1%)
Age	31+	30 (7.8%)
Status	Law student	243 (63.1%)
Status	Law graduate	54 (14.0%)
Status	Advocate or legal professional	53 (13.8%)
Status	Other relevant respondent	35 (9.1%)
Location	Dhaka	149 (38.7%)
Location	District town	87 (22.6%)
Location	Other city corporation	77 (20.0%)
Location	Upazila-rural area	68 (17.7%)
Location	Other	4 (1.0%)
AI training	No	218 (56.6%)
AI training	Yes, short orientation	122 (31.7%)
AI training	Yes, course or certified training	45 (11.7%)

Denominator = 385 for all percentages.

The sample is concentrated in legal education and practice. The questionnaire recorded 195 male respondents, 169 female respondents, eight in the other category and 13 who preferred not to say. These categories describe the sample; the analysis is about gender bias concern rather than estimating gender-group differences. The data do not verify that respondents encountered a biased AI decision.

4.2 Item distributions

Table 3 reports the complete response distributions. Bias concern was endorsed by 289 respondents (75.1%). Testing was endorsed by 215 (55.8%), including 43 in the highest category. Another 138 selected neutrality and 32 disagreed or strongly disagreed. Effective remedy and organisational responsibility were each endorsed by 302 (78.4%), and anticipated confidence from transparency by 292 (75.8%).

Table 3. Distributions of gender bias and accountability ratings

Item	1	2	3	4	5	Agree n (%)	95% CI (%)
GBA 1	0	9	87	179	110	289 (75.1%)	70.5 to 79.1
GBA 2	2	30	138	172	43	215 (55.8%)	50.9 to 60.7
GBA 3	0	15	68	193	109	302 (78.4%)	74.1 to 82.3
GBA 4	0	3	80	198	104	302 (78.4%)	74.1 to 82.3
GBA 5	0	13	80	183	109	292 (75.8%)	71.3 to 79.9

Agree combines 4 and 5. Intervals are Wilson intervals.

The lowest observed endorsement was for organisational testing, but non-endorsement of that item was mostly neutral. In contrast, the effective remedy statement drew 68 neutral responses and 15 disagreements. Equal margins for remedy and organisational responsibility do not imply identical answers from each person. Across the five items, coefficient alpha was 0.270. Pairwise rank correlations ranged from 0.001 to 0.208, with the largest between bias concern and conditional confidence. The pattern and the item content support reporting each judgement separately.

4.3 Joint endorsement and paired contrasts

Table 4 shows that 167 respondents (43.4%) endorsed both testing and remedy. Another 135 endorsed remedy without endorsing testing, 48 endorsed testing without endorsing remedy, and 35 endorsed neither.

Table 4. Joint endorsement of bias testing and remedy

Bias testing	Remedy not endorsed	Remedy endorsed	Total
Not endorsed	35 (9.1%)	135 (35.1%)	170 (44.2%)
Endorsed	48 (12.5%)	167 (43.4%)	215 (55.8%)
Total	83 (21.6%)	302 (78.4%)	385 (100.0%)

Cell percentages use the full sample. Non-endorsement includes neutral responses.

Among testing endorsers, 77.7% also endorsed remedy, compared with 79.4% among those who did not endorse testing. The high remedy endorsement in both groups explains how the two positions can overlap substantially without showing strong discrimination by the testing response. Remedy-only endorsement should not be read as rejection of testing.

Table 5 compares the other four statements with testing. The remedy-testing difference was 22.6 percentage points (95% CI 16.1 to 29.1), arising from 135 versus 48 discordant responses. Organisational responsibility also exceeded testing by 22.6 points. Bias concern exceeded testing by 19.2 points, and conditional confidence exceeded it by 20.0 points.

Table 5. Paired contrasts with bias testing

Contrast	Other only	Testing only	Gap (pp)	95% CI (pp)	Holm p
GBA1 – GBA2	123	49	19.2	12.7 to 25.5	<0.001
GBA3 – GBA2	135	48	22.6	16.1 to 29.1	<0.001
GBA4 – GBA2	134	47	22.6	16.1 to 29.1	<0.001
GBA5 – GBA2	125	48	20.0	13.8 to 26.2	<0.001

Exact McNemar tests. Pointwise percentile intervals use 10,000 respondent resamples.

All four contrasts retained Holm-adjusted p values below 0.001. The evidence concerns relative endorsement as worded. It does not show that respondents would choose remedies over testing if the safeguards were presented as competing investments. Nor does it establish that less specific testing would receive the same responses as a detailed audit proposal.

4.4 Adjusted models

Table 6 reports the two focal terms for each model. In the remedy model, the bias concern coefficient was 0.027 (95% CI –0.078 to 0.132), and the organisational duties coefficient was 0.013 (95% CI –0.125 to 0.151). Neither was supported after adjustment. For conditional confidence, the concern coefficient was 0.204 (95% CI 0.096 to 0.312; Holm-adjusted $p < 0.001$). The duties coefficient was 0.107 (95% CI –0.039 to 0.253; adjusted $p = 0.456$).

Table 6. Adjusted associations with remedy and conditional confidence

Model (n)	Term	Estimate	95% CI	Holm p	R^2
Remedy rating (385)	Concern	0.027	-0.078 to 0.132	1.000	0.029
Remedy rating (385)	Org. duties	0.013	-0.125 to 0.151	1.000	0.029
Conditional confidence (385)	Concern	0.204	0.096 to 0.312	<0.001	0.084
Conditional confidence (385)	Org. duties	0.107	-0.039 to 0.253	0.456	0.084
Remedy endorsement (385)	Concern	1.189	0.873 to 1.620	0.544	0.026
Remedy endorsement (385)	Org. duties	0.952	0.614 to 1.477	0.826	0.026
Law students remedy (243)	Concern	0.048	-0.082 to 0.177	0.472	0.039
Law students remedy (243)	Org. duties	0.122	-0.060 to 0.304	0.380	0.039

Linear rows report unstandardised coefficients, HC3 intervals and ordinary R^2. Binary remedy rows report odds ratios, sandwich intervals and McFadden pseudo-R^2. Holm families: four main coefficients together, two coefficients separately for each sensitivity model.

The models explained 2.9% and 8.4% of observed variance in remedy and conditional confidence ratings, respectively. The concern odds ratio for remedy endorsement was 1.189 (95% CI 0.873 to 1.620). In the law-student analysis the corresponding linear estimate was 0.048 (95% CI –0.082 to 0.177). These remedy sensitivity analyses did not provide clear evidence of a focal association. The coefficients for remedy and conditional confidence were not formally contrasted, so a difference in statistical significance is not a test that their associations differ.

5. Discussion

The study found broad support for remedy and organisational responsibility alongside lower endorsement of the statement about testing. The paired comparison establishes that the difference is visible within the same respondents. A separate result concerns conditional confidence: people who expressed more concern tended to agree more strongly that transparency about testing would increase their confidence. Neither finding establishes how people respond when an organisation actually discloses a test.

5.1 What the testing and remedy difference means

The 22.6-point difference is substantively informative because it separates two governance functions that can be obscured by an average score. Testing identifies possible unequal outcomes at a system level, whereas a remedy is directed at a person who has experienced a biased decision. A person can find the need for relief clear without having a view on how an organisation should test. The remedy item describes an affected person and a desired outcome; the testing item describes a technical and organisational task. Differences in specificity, familiarity and perceived burden are plausible explanations for the response gap. The design does not separate these explanations.

The 167 joint endorsements show that many participants considered both functions important. The 48 testing-only endorsements likewise show that the average gap does not describe everyone. Tests can identify problems that have not yet generated complaints, and accessible review can reveal failures missed by planned tests. A provider that relies only on complaints may miss people who cannot challenge a decision. A provider that relies only on testing may find a group disparity without being able to resolve a particular case. Effective accountability needs a route for each kind of evidence to inform action.[17]

5.2 Concern and anticipated reassurance

Bias concern was positively associated with the rating for anticipated confidence from transparency about bias testing. The result is compatible with the idea that respondents who perceive a risk see greater value in information about how it is addressed. The wording also offers a simpler explanation: both items explicitly refer to bias, and respondents who regard it as salient may agree with both. The study does not observe disclosure, its content or subsequent behaviour. A respondent can expect a disclosure to increase confidence while remaining sceptical about whether a provider's current practice is fair.

A clear report of unfavourable findings may reduce confidence in a tool while improving the quality of a decision about relying on it. This distinction matters for evaluating transparency. Its success should not be defined solely by higher trust scores. An evaluation should ask whether people understand the scope of testing, the limitations and the response to findings. A more informative design would show respondents a specific disclosure and compare comprehension and reliance across conditions.[18]

5.3 Remedy as a procedural commitment

The weak adjusted associations for remedy do not show that bias concern has no relevance to requests for relief. The remedy item is broadly endorsed and asks about people affected by biased decisions, rather than the respondent's personal willingness to complain. There can be broad agreement on the principle while substantial differences remain about the costs, procedures or outcomes people consider acceptable. The interval estimates are compatible with modest associations in either direction. A survey of actual complaints, or a vignette that specifies a forum and the relief it can provide, would answer a different and more practical question.

Procedural justice gives a reason to measure remedy separately from confidence. A person may deserve a meaningful opportunity to challenge a decision whether or not a provider is trusted. If a system is used in a setting where appeal is difficult, public confidence is not a substitute for access to review. The relevant question is whether the institution can receive an account, obtain the evidence and provide an answer or corrective action. The survey indicates support for the principle; it does not show that such a route exists for any

[17] Raji and others (n 3); Bovens (n 3).
[18] Lee and See (n 10); NIST, AI RMF (n 6).

specific system.[19]

5.4 Equality and accountability in Bangladesh

The constitutional equality provisions make the use of AI in public decision processes a relevant subject for domestic legal analysis. The survey does not determine the scope of review or the remedy available in an individual case. A legal inquiry would need to identify the decision maker, the source of authority, the role of the system and the forum able to hear a challenge. It would also need to separate a statistical disparity from the legal assessment of discriminatory treatment. The results are best used to frame questions about what institutions should explain and what affected people should be able to do.[20]

6. Theoretical and Practical Implications

The findings support evaluating prevention, answerability and relief as distinct aspects of accountability. Their relationship is operational: testing can identify a pattern, a review can examine a case, and an authorised institution can change a decision or process. An endorsement score cannot establish that those steps work together.

6.1 Measurement and research design

Future work should measure perceived likelihood and severity of gender-related harm separately from support for particular responses. An audit item should identify who tests, what outcomes are examined and what happens after an adverse finding. A remedy item could describe access to review and the relief that the reviewer can offer. Confidence should identify the actor or system in question. These distinctions allow an empirical model to evaluate whether concern relates differently to confidence, willingness to use a service and demand for review. They also make clear which measures describe attitudes rather than performance.[21]

6.2 Institutional design

Organisations considering a consequential tool can specify the decisions it supports and review whether the proposed evaluation covers the people and conditions of intended use. An audit plan can record who owns each check, what evidence is collected and how a finding can delay or change deployment. Reporting should identify outcomes, comparison groups and material limitations without exposing respondents or assuming that every relevant category is captured. If the tool changes, the institution should know which checks need to be repeated. A transparent report has practical value when the institution can show how the findings inform decisions.[22]

[19] Bovens (n 3); Culnan and Armstrong (n 9).
[20] Constitution of Bangladesh; Selbst and others (n 1).
[21] Selbst and others (n 1); Lee and See (n 10).
[22] Raji and others (n 3); NIST, AI RMF (n 6).

An accessible remedy process can begin with a clear way to raise a concern even when a person is unsure how AI was involved. The reviewer needs access to the relevant records and authority to correct or escalate the decision. A user-facing explanation should identify the next step and what evidence can be considered. A process evaluation could test whether people with different language, access and skill needs can locate the channel and obtain a reasoned response. Complaints can inform monitoring, but a low number of complaints is not evidence that a system is fair or that a channel is easy to use.[23]

6.3 Policy and consultation

Consultation could describe an actual or proposed use and ask separately about the scope of testing, disclosure of findings and access to review. This would help distinguish uncertainty about a technical practice from opposition to it. Participants should be able to express that they favour use only in certain settings or that they want remedies regardless of whether a tool is ultimately adopted. The present results motivate these questions; they do not identify the optimal regulatory instrument or show that one form of oversight will work across sectors.

7. Conclusion Limitations and Future Research

Among 385 adult respondents, effective remedy and organisational responsibility received greater endorsement than testing for gender-related discriminatory outcomes. Forty-three per cent endorsed both remedy and testing. Bias concern was associated with anticipated confidence from transparency about testing, while adjusted evidence for remedy ratings was weak. The distinction between preventive testing, access to relief and expected reassurance is the central contribution. These outcomes should be measured and evaluated separately when institutions consult on AI governance.

The survey includes substantial representation from legal education and practice, and the recruitment frame, invitation denominator and dates were not documented in the materials. The results cannot be generalised to all people in Bangladesh. Single statements differ in wording and concreteness. The analysis does not observe discriminatory outcomes, use of a remedy or the effect of disclosing an audit. Cross-sectional associations can reflect unmeasured background factors or common item wording. Other articles based on this common Group 10 instrument share its records and do not constitute independent samples.

A follow-up could recruit people with experience of specified automated decisions and document recruitment across relevant service settings. Cognitive interviews could examine what participants understand by

[23] Bovens (n 3).

testing, discrimination and an effective remedy. A factorial study could vary whether a provider conducts testing, discloses its scope and offers independent review while holding the service and decision constant. It could record comprehension, perceived fairness, willingness to use the service and intended challenge separately. An implementation study could then determine whether review processes identify failures and lead to corrective action. These designs would connect the present study's stated judgements to measurable institutional outcomes.

Declarations

Ethical approval and oversight

No study-specific formal ethics approval or exemption was documented in the available materials. Section 3.6 describes the recorded consent, exclusion of contextual text and aggregate reporting.

Informed consent

All 385 analysed records contain affirmative consent and adult eligibility responses.

Data availability

Requests for de-identified numerical data and analytical specifications may be directed to the corresponding author, subject to confidentiality and applicable institutional requirements.

Funding and competing interests

No external funding or competing interest was disclosed in the available study materials.

Author contributions

Surove Syed Oaishe led the student research, developed the research question, organised the study materials and prepared the initial manuscript. Prof. Dr Kazi Abdul Mannan supervised the research, provided conceptual and methodological guidance, contributed to interpretation and critically revised the manuscript. Kazi Abdul Mannan is the corresponding author.

Generative AI disclosure

Generative AI assisted literature retrieval, drafting, editing and preparation of analytical code. Numerical estimates were computed programmatically from the supplied workbook. The authors are responsible for verifying and approving the analysis, interpretation and reporting.

Bibliography

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

Buolamwini J and Gebru T, ‘Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification’ (2018) 81 Proceedings of Machine Learning Research 77

Culnan MJ and Armstrong PK, ‘Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation’ (1999) 10 Organization Science 104

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

Holm S, ‘A Simple Sequentially Rejective Multiple Test Procedure’ (1979) 6

Scandinavian Journal of Statistics 65

Hosmer DW Jr, Lemeshow S and Sturdivant RX, *Applied Logistic Regression* (3rd edn, John Wiley & Sons 2013)

Lee JD and See KA, 'Trust in Automation: Designing for Appropriate Reliance' (2004) 46 Human Factors 50

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

National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1, 2023) doi:10.6028/NIST.AI.100-1

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

Selbst AD and others, 'Fairness and Abstraction in Sociotechnical Systems' (Proceedings of the Conference on Fairness, Accountability, and Transparency, ACM 2019) 59

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

von Elm E and others, 'The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies' (2007) 4 PLOS Medicine e296

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

White H, 'A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity' (1980) 48 Econometrica 817

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

Appendix A Questionnaire wording

Table A1 reproduces the focal accountability statements and the items used for the AI orientation adjustment.

Table A1. Exact wording of the analysed questionnaire statements

Code	Questionnaire statement
AIL1	I understand that generative AI can produce plausible but inaccurate information.
AIL2	I verify important AI-generated information before relying on it.
AIL3	I understand that AI outputs may reflect bias or incomplete data.
GBA1	I am concerned that AI systems can reproduce gender bias present in data or institutions.
GBA2	Organizations should test AI systems for gender-related discriminatory outcomes.
GBA3	People affected by biased AI decisions should have an effective remedy.
GBA4	Developers and deploying organizations should be accountable for preventable discriminatory harms.
GBA5	Transparency about bias testing would increase my confidence in AI systems.

Source: Group 10 Codebook. Response anchors run from 1 strongly disagree to 5 strongly agree.

GBA3 and GBA5 are separate outcomes. The organisational duties mean contains GBA2 and GBA4 only.

Table of Legislation

Constitution of the People’s Republic of Bangladesh, arts 27-28