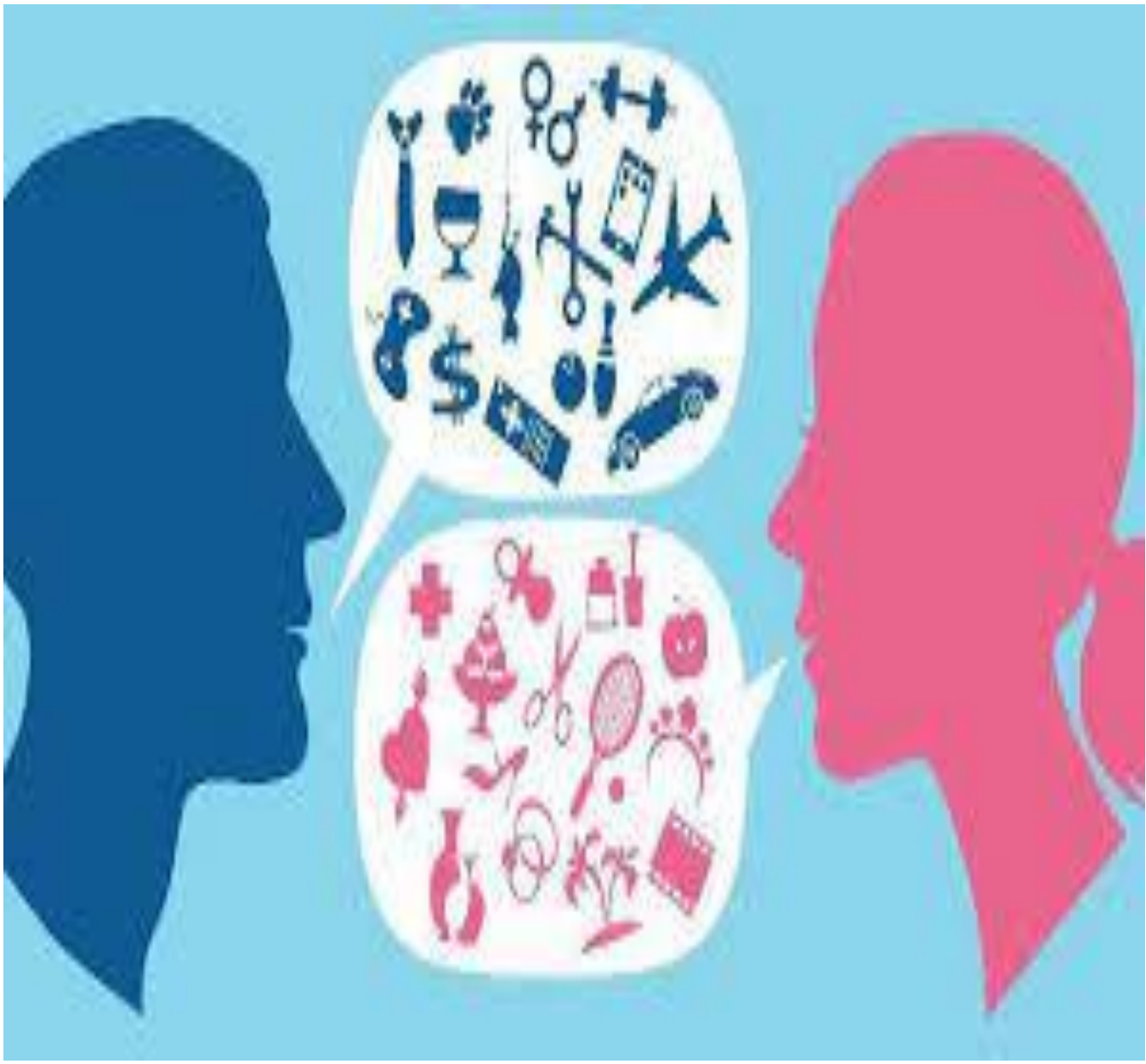
Journal of Gender Education and Society

Volume: 5 Issue: 4
(October-December), 2026
ISSN: 2957-4137



KMF PUBLISHERS www.kmf-publishers.com/jges/

Journal of Gender Education and Society

Volume 5 Issue 4, Year 2026

Research Article

OPEN

DOI: https://doi.org/10.64907/xkmf.v5.i4.jges.1

Gender Differences in Conditional Willingness and Trust Expectations for AI Legal Services in Bangladesh

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

Artificial intelligence may widen access to legal information while reproducing inequalities in who can use, evaluate and challenge its outputs. Research on gender and legal AI therefore needs to distinguish conditional willingness to use a service from confidence that its design supports trustworthy operation. This study examines that distinction in the supplied Group 10 Law-SDG database, linking gender and technology, procedural justice and digital inclusion perspectives. All 385 records inform descriptive analysis; comparisons between respondents recorded as female and male use 364 records, with the remaining gender responses retained descriptively. Item-level analysis combines distributional comparisons, stratified bootstrap intervals, paired contrasts and covariate-adjusted logistic models. Conditional willingness was endorsed by 307 records (79.7%), whereas 198 (51.4%) endorsed the proposition that inclusive design can improve trust. Their paired difference was 28.3 percentage points (95% bootstrap interval 21.8-34.8). The female-minus-male willingness difference was -0.7 points (-9.0-7.5); the adjusted difference was 1.0 point (-7.4-9.4). Inclusive-design trust expectations and discrimination salience likewise showed small estimated gender contrasts with substantial uncertainty. Low block reliability precluded latent-scale modelling. The findings support neither a gender-deficit explanation nor a claim of established equality: similar reported willingness can coexist with unequal access and unmeasured exposure to harm. The contribution is a measurement-conscious distinction between conditional acceptance, trust expectations and substantive inclusion. Interpretation remains bounded by undocumented recruitment and by codebook labels identifying respondent sequences and comments as synthetic; numeric response provenance requires verification before these records can support claims about a human population.

Keywords: legal artificial intelligence; gender; conditional willingness; procedural justice; digital inclusion; Bangladesh; SDG 5; SDG 16

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

Corresponding author: Md. Didarul Islam. Email: dihanalfididar@gmail.com

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/).

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

1. Introduction

Artificial intelligence in legal services raises a question that adoption statistics alone cannot answer: under what conditions is use compatible with equal access, informed reliance and an effective opportunity to challenge error? A person may welcome convenient legal information while remaining uncertain about the institution supplying it. Another may value inclusive design yet be unable to access a service privately or affordably. Understanding gender differences requires attention to these distinctions, especially where the evidence consists of statements about hypothetical services rather than records of completed legal transactions.

The Bangladesh setting makes this question relevant to an active governance conversation. UNESCO's account of the national AI readiness validation process identifies gender and access gaps alongside institutional capacity as matters for attention. That account supplies policy context, rather than evidence about the present respondents or proof of the effects of any particular legal AI service. It also suggests why an analysis of acceptance should ask how services become accessible and accountable, rather than treat greater willingness as a sufficient measure of success.[1]

The sustainable development connection is similarly substantive. SDG 5 concerns gender equality, including enabling technology, while SDG 16 includes equal access to justice and inclusive institutions. An AI interface may contribute to these aims when it helps people understand rights and reach competent assistance. The same interface could undermine them if inaccurate guidance, unaffordable access or inaccessible complaints procedures prevent meaningful use. Survey attitudes are therefore an intermediate source of evidence, not a direct indicator that an SDG target has been achieved.[2]

Legal information also differs from many consumer technology applications. A plausible answer can influence whether someone preserves evidence, seeks advice or pursues a remedy. Research demonstrating legal hallucinations in large language models makes the distinction between fluent output and dependable assistance especially important. The present study does not test the accuracy of any system, and findings from other jurisdictions cannot quantify error in Bangladesh. They do establish a reason to take the safeguards embedded in the willingness question seriously.[3]

Gender and technology scholarship provides a starting point for examining variation without presuming a universal hierarchy of acceptance. Earlier research investigated how gender could condition the relationship between social influence, perceptions and technology use. Its relevance lies in treating acceptance as situated, rather than assuming every group responds to the same institutional environment in the same way. Its findings are not transferred as fixed predictions about Bangladeshi women and men or about legal services that were not studied in that research.[4]

The supplied research plan identifies Md. Didarul Islam's topic as gender differences

[1] UNESCO, 'AI for All: Shaping Bangladesh's Future with Ethical AI' (22 May 2025) <https://www.unesco.org/en/articles/ai-all-shaping-bangladeshs-future-ethical-ai> accessed 25 October 2026.

[2] UN General Assembly, 'Transforming Our World: The 2030 Agenda for Sustainable Development' (25 September 2015) UN Doc A/RES/70/1.

[3] Matthew Dahl and others, 'Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models' (2024) 16 Journal of Legal Analysis 64.

[4] Viswanath Venkatesh and Michael G Morris, 'Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior' (2000) 24 MIS Quarterly 115.

in trust and adoption of AI legal services. Close examination of the questionnaire requires a more precise empirical formulation. GLA2 asks whether a respondent would use an AI legal service if it met their needs and appropriate safeguards existed. GLA4 asks whether inclusive design can improve trust. Neither statement measures actual service adoption, verified safeguards, present trust in a named institution or experience of successful legal assistance. The article therefore studies conditional willingness and inclusive-design trust expectations while preserving the substantive concern of the planned topic.

This distinction provides the article's first contribution. Combining these statements into a single acceptance score would conceal their different objects: a personal conditional intention, a belief about design and trust, a concern about discrimination, an equality norm and support for outcome testing. Item-level analysis protects the connection between a question and the conclusion it can sustain. This is particularly necessary because the supplied blocks have low internal consistency, making a reflective trust or adoption interpretation difficult to defend.

The second contribution concerns the interpretation of gender comparisons. A small estimate accompanied by a broad interval does not establish equality. Conversely, an unsupported expectation of a large gender gap should not be rescued by searching across numerous outcomes. The analysis therefore reports estimates, uncertainty and multiplicity-adjusted comparisons, then asks what magnitudes remain compatible with the data. This approach treats an inconclusive contrast as information about precision rather than a reason to describe the study as either a success or a failure.[5]

The third contribution is a paired comparison within records. The difference between conditional willingness and endorsement of the inclusive-design trust proposition may be larger than differences between recorded gender categories. Such a result would redirect attention from presumed group deficits toward the institutional conditions under which willingness is expressed. Because the two items ask different questions, however, their difference is a response-pattern contrast, not a calibrated psychological distance or a measure of a trust deficit.

The article addresses three research questions. First, how do the distributions of conditional willingness, inclusive-design trust expectations and discrimination salience differ between female and male records? Second, do these contrasts change after adjustment for recorded age, legal status, use frequency, training and rural location? Third, how large is the within-record willingness-trust-expectation contrast, and how precisely can its gender variation be estimated? Smaller recorded gender categories remain visible descriptively but are not used to support unstable regression claims.

These questions define a contribution distinct from the other Group 10 topics. The analysis does not centre perceived bias as a predictor of algorithmic accountability, nor AI literacy as a predictor of women's legal empowerment. It instead evaluates the measurement and precision of gender comparisons in conditional service acceptance. The scope of all conclusions is the supplied database: recruitment details are unavailable, and synthetic labels in the

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

codebook require explicit provenance qualification before any population interpretation.

2. Literature Review and Theoretical Foundation

This section connects three perspectives around a common problem: expressed acceptance can depend on social position, perceived procedural conditions and practical access. The theories guide the choice and interpretation of observable items. They are not presented as latent constructs that the questionnaire has already validated. The resulting framework distinguishes between conditional willingness, a proposition about design and trust, and the salience of discrimination concerns.

2.1 Gender, technology and contextual acceptance

Gender-sensitive technology research asks whether relationships among expectations, social influences and use vary across contexts and groups. The unified theory of acceptance and use of technology integrates several established explanations, including performance expectations, effort expectations, social influence and facilitating conditions. This provides a vocabulary for considering why willingness may depend on circumstances rather than on a demographic characteristic alone. The present questionnaire does not contain a validated battery for all these constructs, so the study does not estimate or claim to test that full model.[6]

A feminist technology perspective adds an institutional question: whose needs and practices are taken as the default when technology is developed? Faulkner's analysis locates gender in relationships among technologies, practices and social organisation. Applied here, this means that an apparent gender difference could reflect access conditions or service design rather than an inherent propensity toward technology. It also means that an apparently similar response distribution need not imply a common experience of the service environment.[7]

The difference between these propositions matters for interpretation. If a study finds lower willingness among women, training is only one possible response; the problem might instead involve confidentiality, inaccessible interfaces or credible concern about discrimination. If a study finds similar willingness, barriers may still emerge when respondents attempt to act on that willingness. The empirical task is therefore to estimate a contrast accurately, then resist attaching mechanisms that were never measured. Gender is treated as a reported social category, not a randomised exposure or a biological explanation.

The database records male, female, other and prefer-not-to-say responses. Its categories do not describe gender identity in sufficient detail to support claims about particular minority communities. Treating the other category as a homogeneous substantive group would exceed its wording. Treating non-disclosure as a gender identity would do the same. Descriptive retention, clear denominators and restraint in inference provide a defensible response to this measurement limitation.

[6] Viswanath Venkatesh and others, 'User Acceptance of Information Technology: Toward a Unified View' (2003) 27 MIS Quarterly 425.

[7] Wendy Faulkner, 'The Technology Question in Feminism: A View from Feminist Technology Studies' (2001) 24 Women's Studies International Forum 79.

2.2 Conditional willingness, adoption and trust

Adoption requires a behavioural reference: an identifiable service, an opportunity to use it and evidence of uptake. Conditional willingness is further upstream. General theories of behavioural intention distinguish intention from behaviour, with actual performance depending on circumstances and perceived control. GLA2 is especially conditional because it explicitly supplies need fulfilment and appropriate safeguards as premises. Its endorsement is evidence of willingness under those premises, not willingness to use every available tool.[8]

Trust is also more specific than a favourable opinion. Organisational trust scholarship connects trust to vulnerability and the characteristics of a trustee; related interdisciplinary work distinguishes trust from its antecedents and consequences. In a legal AI setting, the relevant trustee could be a developer, a legal aid provider, a professional or a public authority. A general statement about inclusive design does not identify which of these actors is trusted or what reliance is accepted.[9]

GLA4 is consequently interpreted as an inclusive-design trust expectation: agreement that inclusive design can improve trust. It does not establish that a current service is inclusive, that the respondent presently trusts one, or that an intervention would increase trust by a measurable amount. Formal work on trust in AI likewise underscores the need to specify what the person relies on the system to do. This supports keeping the two focal items separate rather than assuming that their similar subject matter makes them interchangeable indicators.[10]

Studies of public-sector chatbots and explanations of automated decisions illustrate that institutional information and explanations can matter for perceived trustworthiness. They do not establish that the present inclusive-design item captures the same construct or that an observed association would identify the effect of design. Their relevance is to motivate more specific institutional questions in future research, including who provides the service, how errors are corrected and what an explanation enables a user to do.[11]

2.3 Procedural justice and discrimination salience

Procedural justice redirects attention from whether a tool is convenient to whether the processes surrounding its use are experienced as fair. Tyler's work links perceived legitimacy and fair treatment to relations with legal authorities. For AI legal services, a procedural interpretation invites questions about respectful interaction, opportunities to communicate relevant facts and the possibility of contesting an adverse answer. The present survey does not directly measure these dimensions or legitimacy; procedural justice is an interpretive foundation and a guide for future instrument development.[12]

[8] Icek Ajzen, 'The Theory of Planned Behavior' (1991) 50 Organizational Behavior and Human Decision Processes 179.

[9] 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.

[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.

[11] Naomi Aoki, 'An Experimental Study of Public Trust in AI Chatbots in the Public Sector' (2020) 37 Government Information Quarterly 101490; Stephan Grimmelikhuijsen, 'Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making' (2023) 83 Public Administration Review 241.

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

GLA3 states that concerns about discrimination affect willingness to use AI legal services. The direction is not specified. Endorsement could indicate avoidance, selective use or a preference for services accompanied by safeguards. It could also reflect concern about discrimination in existing legal institutions. Coding this item as fear, resistance or low acceptance would therefore impose meaning that the wording does not supply. It is analysed as discrimination salience and is not reverse-scored into an adoption index.

There is a substantive basis for taking concern seriously without attributing unobserved experiences to respondents. Gender Shades demonstrated intersectional performance differences in commercial facial analysis, illustrating why aggregate performance can obscure unequal outcomes. It studied a different technology and setting, so its error rates are not extrapolated to legal AI. The analytical lesson is that system performance and user attitudes require separate investigation, including attention to intersecting social positions.[13]

Intersectional legal scholarship further cautions that a single-axis comparison can leave important forms of disadvantage analytically invisible. This study's female-male contrast is a limited descriptive comparison, not a complete account of gender justice. The database lacks sufficient information and subgroup sizes to examine how disability, income, minority status or household circumstances intersect with gender. The response is to state that boundary and specify better future measurement, rather than infer the absence of compounded disadvantage.[14]

2.4 Digital inclusion and access to justice

Digital inclusion concerns whether people can convert technological resources into meaningful participation. Helsper's corresponding-fields model relates digital and social exclusion across domains, encouraging analysis beyond device ownership alone. In this study, recorded frequency and training can describe aspects of familiarity, while location can identify a broad contextual distinction. None directly measures reliable connectivity, private access, affordability, language accessibility or the ability to obtain a remedy after harmful advice.[15]

This creates a practical distinction between willingness equality and access equality. Two respondents may both endorse GLA2 because both imagine a safe service. One may nonetheless lack private access to a device, time to use the service or confidence that a legal professional will help when the output is wrong. A survey that supplies safeguards in its question cannot establish whether these conditions exist. Digital inclusion therefore explains why positive conditional willingness is a potential resource for service development but an incomplete measure of opportunity.

AI literacy may contribute to informed participation by helping users recognise limitations and evaluate outputs. Long and Magerko organise competencies and design considerations relevant to such participation. The common questionnaire contains several

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

[14] Kimberlé Crenshaw, 'Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics' [1989] University of Chicago Legal Forum 139.

[15] Ellen Johanna Helsper, 'A Corresponding Fields Model for the Links Between Social and Digital Exclusion' (2012) 22 Communication Theory 403.

related statements, but they mix awareness, self-reported verification, confidentiality practice and oversight preferences. Their inclusion does not establish an objective literacy score. The present article therefore avoids making the literacy block its central mechanism or reproducing the separate women's empowerment topic.[16]

2.5 Framework and analytical expectations

Figure 1 sets out the conceptual distinction. Gender is examined as a grouping variable within recorded social and technological circumstances. Conditional willingness, inclusive-design trust expectations and discrimination salience are separate responses. The line connecting design expectations and willingness represents an exploratory association, not a causal sequence. Structural access conditions and actual legal outcomes remain outside the instrument's coverage.

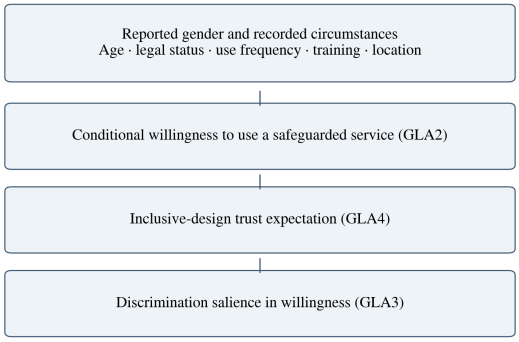


Figure 1. Conceptual scope: separate responses examined within recorded social and technological circumstances.

Lines indicate analytical grouping, not causal effects. Structural access conditions and actual legal outcomes are not measured. The framework yields three explicitly exploratory expectations. E1 is that female and male response distributions may differ on conditional willingness, trust expectations or discrimination salience; the direction is left open because the theories do not warrant a universal prediction in this setting. E2 is that adjustment for measured circumstances may change these contrasts, without necessarily identifying or removing the processes through which inequality operates. E3 is that the willingness-trust-expectation contrast may vary by gender, while an association between inclusive-design beliefs and willingness may also differ across the two groups.

These expectations were formulated for this secondary analysis after the database became available and were not preregistered. They are not relabelled as confirmed hypotheses after inspecting results. Three-item comparison families receive multiplicity adjustment. The exploratory interaction is reported with its interval and raw p-value, and practical interpretation rests on effect size, measurement and precision. No prediction is made that women should have lower willingness or that a non-significant comparison establishes parity.

A final methodological implication follows from equivalence research. Demonstrating that a difference is sufficiently small requires a substantively justified margin, rather than a failure to reject zero. The supplied plan specifies no such margin. The article therefore reports a sensitivity analysis across illustrative margins, clearly distinguished from a confirmatory equivalence claim. This makes the available precision visible while leaving the normative

[16] Duri Long and Brian Magerko, 'What Is AI Literacy? Competencies and Design Considerations' in Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (ACM 2020).

decision about an acceptable service disparity to an independently justified research and policy process.[17]

3. Methodology

The study is a secondary, cross-sectional analysis of the supplied Group 10 workbook, interpreted alongside the common questionnaire and the research-plan entry. The design prioritises fidelity to item wording, transparent denominators and uncertainty estimates. It does not recreate undocumented field procedures or treat a recommended research protocol as proof that those procedures occurred.

3.1 Materials, provenance and analytical population

The workbook contains a Data sheet with 385 rows and 35 variables, together with README and Codebook sheets. The accompanying questionnaire provides the response wording and consent information. The research plan proposes a survey of 385 and describes probability sampling as preferred. The available materials do not document a sampling frame, recruitment dates, selection probabilities, contact attempts, institutional clusters or a response rate. Those planned features are therefore not stated as completed methods.

A material provenance qualification applies. The codebook describes Respondent_ID as a synthetic respondent sequence and the optional comment as a synthetic contextual comment. These labels do not by themselves establish whether every numeric response is simulated, but they prevent treating the workbook as verified human-participant evidence without further documentation. Comments are excluded entirely. All numerical results below describe the supplied records; inferential intervals are conditional analytical summaries under an independent-record model, not proof of population representation.

All 385 records meet the workbook's recorded adult eligibility, awareness and consent conditions. All enter descriptive analysis and the paired item comparison. Female-male analyses use 169 female and 195 male records, giving n = 364. Eight other and 13 prefer-not-to-say records remain in descriptive tables. Their exclusion from the two-category model is a limitation of that specific comparison, not exclusion from the study database, and no pooled binary gender recoding is performed.

The plan's Cochran calculation corresponds to approximately 385 observations for a large population under conventional simple-random-sampling assumptions. Matching that number does not establish those assumptions or provide a five-per-cent margin of error for this workbook. No survey weights or finite population correction are applied because the required design information is unavailable. No retrospective power calculation is offered as evidence of adequacy; the observed confidence intervals communicate precision more directly.[18]

[17] Daniël Lakens, 'Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses' (2017) 8 Social Psychological and Personality Science 355.

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

3.2 Variables and measurement decisions

Table 1 states the analytical role of each focal item. Scores range from one, strongly disagree, to five, strongly agree. Endorsement is defined as four or five; neutrality and disagreement are combined only for that binary summary. The full distributions remain visible, and rank comparisons preserve the original ordering. No item is interpreted as an objective test of legal knowledge, actual exposure to discrimination or completed service use.

Table 1. Focal items and analytical interpretation

Code	Abbreviated content	Interpretation
GLA1	Equal access across genders	Normative equality statement
GLA2	Would use if needs and safeguards met	Primary conditional-willingness outcome
GLA3	Discrimination concerns affect willingness	Discrimination salience; direction unspecified
GLA4	Inclusive design can improve trust	Trust expectation; not current trust
GLA5	Testing for unequal group outcomes	Outcome-testing preference

Source: supplied questionnaire and codebook. Appendix A gives exact wording. The statements are not assumed to form a reflective scale.

Internal consistency was examined for the four supplied five-item blocks. Cronbach's alpha was 0.215 for AIL, 0.247 for GLA, 0.270 for GBA and 0.288 for WLE. These values offer no basis for treating the blocks as reliable reflective scales in this analysis. Alpha is not itself a complete validity test, and conceptually heterogeneous items may appropriately remain distinct. Accordingly, no composite trust, adoption, literacy or empowerment score, structural equation model or mediation model is estimated.[19]

3.3 Audit and descriptive analysis

The audit checked identifier uniqueness, missingness, response ranges, repeated substantive vectors and uniform responding across the 20 agreement items. It found no missing or out-of-range substantive responses, duplicate identifiers, duplicate 20-item vectors or complete straightliners. Fifteen tool-name responses are missing; tool name is not used in modelling. These checks establish technical consistency of selected fields, not authenticity, respondent attentiveness or the validity of recruitment.

The two experience fields are discordant: 15 records report never having used AI while recording a non-never frequency, and nine record never frequency with non-never experience. No record reports never in both fields. The primary analysis retains them because the data do not establish which field is correct. A sensitivity analysis excludes either never response within the female-male subset. Optional comments cannot resolve these inconsistencies because the codebook explicitly labels them synthetic.

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

3.4 Estimation, multiplicity and sensitivity

Female-male distributional contrasts use two-sided Mann-Whitney tests with tie correction, interpreted as tests of distributions rather than automatically as tests of medians. Probability of superiority is the chance that a randomly selected female record has a higher response than a male record, with half credit for ties. Three tests covering GLA2, GLA4 and GLA3 form one Holm-adjusted family. Endorsement differences are female minus male, with 10,000 gender-stratified percentile bootstrap resamples and pointwise 95% intervals.[20]

Separate logistic models estimate endorsement for each item. Each includes gender, four age indicators, current law-student status, ordered AI frequency, any formal training and upazila-rural location. Frequency enters as one linear step across five categories; other contextual variables use the stated binary contrasts. The specification is parsimonious and does not exhaust available social circumstances. All models include an intercept. They describe conditional associations, because the cross-sectional design and unspecified selection process do not identify causal gender effects.

Model intervals use an HC1 sandwich covariance with a sample-size correction. Standardised endorsement differences average predictions over the same 364 recorded covariate profiles, setting the gender indicator first to male and then female. Delta-method intervals use the model covariance and hold that empirical covariate distribution fixed. They are descriptive contrasts under a fitted model, not estimates of changing an individual's gender. Three gender-coefficient p-values form a second Holm family.[21]

The paired analysis subtracts GLA4 endorsement from GLA2 endorsement within every record and bootstraps whole records to preserve dependence. An exact McNemar test uses the discordant responses. The female-male difference in these paired contrasts uses stratified resampling. A separate exploratory model adds centred GLA4 and its interaction with gender to the GLA2 model; only its coefficients, not a causal mediation pathway, are interpreted.[22]

Sensitivity analyses retain the original one-to-five outcomes in linear regressions with HC3 covariance, repeat the willingness model among consistent-use records and among current law students, and compare pointwise 90% bootstrap intervals against illustrative symmetric bounds of five and ten percentage points. These bounds were not set prospectively or justified as acceptable inequalities. The exercise concerns sensitivity of interpretation, not a formal declaration of gender equivalence.[23]

Calculations use Python, pandas, NumPy and SciPy. Logistic maximum likelihood is independently checked against an unpenalised-limit scikit-learn fit; maximum absolute coefficient discrepancies for the three primary models are below 0.000001. A fixed random seed of 20261025 makes resampling reproducible. Appendix B records the coding, estimands and fitting checks, and Appendix C reports the complete adjusted coefficients.

[20] Henry B Mann and Donald R Whitney, 'On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other' (1947) 18 Annals of Mathematical Statistics 50; 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).

[21] Leonard A Stefanski and Dennis D Boos, 'The Calculus of M-Estimation' (2002) 56 American Statistician 29.

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

[23] James G MacKinnon and Halbert White, 'Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties' (1985) 29 Journal of Econometrics 305.

3.5 Ethics and responsible reporting

No formal ethics approval or documented exemption accompanies these materials, and none is claimed. The supplied participant information specifies adult participation, voluntariness, the ability to stop, avoidance of names and reporting in aggregate. The workbook records consent for all rows. These documents support an account of the intended protections; they do not independently verify their implementation or establish that institutional review was unnecessary. The absence of an approval route cannot itself substitute for an ethical determination.

The present analysis applies concrete safeguards within its control: direct identifiers are not introduced, synthetic comments are excluded, no individual is profiled, small gender categories are reported only in aggregate, and findings are not framed as inherent group deficiencies. Data provenance, measurement limits and missing governance documentation are disclosed rather than obscured. A substantiated final human-participant submission requires the data custodian to document recruitment, consent implementation, access permissions and the applicable ethics-review position. Ethical credibility depends on that factual record as well as on careful secondary analysis.

4. Results

The results move from database coverage to item distributions, gender contrasts and adjusted estimates. The central numerical distinction is between relatively high conditional willingness and more moderate endorsement of the inclusive-design trust proposition. The gender comparisons are reported with intervals so that small point estimates are not mistaken for established equality.

4.1 Database profile and coverage

Table 2 describes all 385 records. The age distribution is concentrated in the 21-23 category, and current law students constitute 243 records. This composition matters because a legal-education-related sample can express different expectations from people seeking urgent legal assistance. The workbook includes graduates, legal professionals and other relevant respondents, so neither the whole database nor the inferential subset is described as consisting entirely of students.

Table 2. Profile of the complete database (N = 385)

Characteristic	Recorded categories and counts
Age	21-23: 177; 24-26: 83
	18-20: 64; 27-30: 31
	31+: 30
Gender	Male: 195; Female: 169
	Prefer not to say: 13; Other: 8
Current status	Law student: 243; Law graduate: 54
	Advocate or legal professional: 53; Other relevant respondent: 35
Education	Undergraduate: 220; Graduate: 77
	Postgraduate: 45; Professional: 43
Location	Dhaka: 149; District town: 87
	Other city corporation: 77; Upazila-rural area: 68
	Other: 4
AI frequency	Weekly: 160; Daily or almost daily: 101
	Monthly: 74; Less than monthly: 41
	Never: 9
Training	No: 218; Yes, short orientation: 122
	Yes, course or certified training: 45

Counts are unweighted and sum to 385 within each characteristic. Current status and education level are different variables. The other relevant respondent category is not further specified.

Female and male categories account for 364 records, or 94.5% of the database. The remaining 21 records are retained in the

whole-database figures and paired analysis. Location categories indicate some geographical variation, including 68 upazila-rural records, but they do not establish national coverage or probability sampling. Training and frequency likewise describe the records without establishing competence, access quality or the suitability of AI for legal tasks.

4.2 Item distributions and conditional acceptance

Table 3 presents the five GLA items separately. Conditional willingness is endorsed in 307 records, compared with 203 for equal access, 212 for discrimination salience, 198 for inclusive-design trust expectations and 193 for outcome testing. These differences in endorsement are analytically important because a common topic heading could otherwise imply that all five statements describe one homogeneous favourable attitude toward AI legal services.

Table 3. Full item distributions and endorsement (N = 385)

Item	Mean (SD)	Counts for scores 1 / 2 / 3 / 4 / 5	Endorsed n (%)
GLA1: Equal access	3.56 (0.86)	2 / 37 / 143 / 150 / 53	203 (52.7%)
GLA2: Conditional willingness	4.05 (0.74)	0 / 9 / 69 / 200 / 107	307 (79.7%)
GLA3: Discrimination salience	3.57 (0.81)	2 / 30 / 141 / 169 / 43	212 (55.1%)
GLA4: Inclusive-design trust expectation	3.57 (0.81)	1 / 27 / 159 / 148 / 50	198 (51.4%)
GLA5: Outcome testing	3.53 (0.81)	1 / 30 / 161 / 149 / 44	193 (50.1%)

Endorsed means four or five. Score three is neutral. Means are descriptive score summaries and do not establish equal-interval measurement or scale validity.

No record strongly disagrees with GLA2, and 107 strongly agree. For GLA4, 159 select the neutral category, compared with 69 for GLA2. Neutrality is not coded as distrust in the descriptive interpretation: the item asks whether a particular design approach can improve trust, and a neutral response can reflect uncertainty about that proposition. Similarly, respondents who do not endorse a general testing statement cannot automatically be described as opposing equality or approving discriminatory systems.

The observed distributions are consistent with broad conditional openness alongside less settled agreement about the particular role of inclusive design. They do not demonstrate a successful legal service, a reduction in unequal access or a response to an implemented intervention. Those outcomes were not measured. The results should consequently be read as a structured description of questionnaire responses, with the provenance boundary applying to every percentage and model.

4.3 Gender contrasts in distributions and endorsement

Table 4 displays the three focal outcomes across all recorded gender categories. Female willingness endorsement is 134/169 (79.3%), and male endorsement is 156/195 (80.0%). Inclusive-design trust-expectation endorsement is 50.9% and 50.3%, respectively. Discrimination-salience endorsement is 54.4% and 54.4% after rounding. The smaller categories are shown with exact denominators, making their limited basis visible without creating a pooled residual gender group.

Table 4. Focal endorsement by recorded gender category

Gender	n	Willingness n (%)	Trust expectation n (%)	Discrimination salience n (%)
Female	169	134 (79.3%)	86 (50.9%)	92 (54.4%)
Male	195	156 (80.0%)	98 (50.3%)	106 (54.4%)
Other	8	7 (87.5%)	6 (75.0%)	6 (75.0%)
Prefer not to say	13	10 (76.9%)	8 (61.5%)	8 (61.5%)

All 385 records are represented. The two-category comparison uses 364 female or male records. Other and non-disclosing responses are retained descriptively; no unstable small-group model is fitted.

The female-minus-male willingness difference is -0.7 percentage points, with a pointwise 95% bootstrap interval of -9.0 to 7.5. For inclusive-design trust expectations, the difference is 0.6 points (-9.2 to 11.1); for discrimination salience, 0.1 points (-10.3 to 10.4). Thus the point estimates are close to zero, while differences of several points in either direction remain compatible with the resampling intervals. None of these intervals corrects the uncertainty introduced by unknown recruitment or data provenance.

Table 5 complements endorsement summaries with distributional comparisons. Probability-of-superiority estimates are 0.481 for willingness, 0.505 for trust expectations and 0.511 for discrimination salience. Values close to one-half indicate little overall ordering between the two recorded distributions. All three Holm-adjusted distributional p-values equal 1.000 after truncation at one. These results provide no compelling evidence of a distributional gender difference in this dataset; they are not a test that all differences relevant to legal access are absent.

Table 5. Female-minus-male endorsement and rank comparisons (n = 364)

Outcome	Gap pp	95% bootstrap interval	Superiority	Raw rank p	Holm p
Conditional willingness	-0.7	-9.0-7.5	0.481	0.493	1.000
Trust expectation	0.6	-9.2-11.1	0.505	0.864	1.000
Discrimination salience	0.1	-10.3-10.4	0.511	0.686	1.000

Superiority is P(female score > male score) + 0.5 P(tie). Rank tests retain all five categories. Bootstrap intervals concern endorsement differences and are pointwise; the Holm family contains the three rank tests.

The contrast between a small observed difference and a broad compatible range is especially relevant for service decisions. A nine-point difference might matter for a large programme even when the point estimate is below one point. The appropriate conclusion is therefore specific: the database does not reveal a clear female-male contrast on these statements, and its precision is insufficient to dismiss every modest disparity. Neither a deficit narrative nor a parity narrative follows from the p-values alone.

4.4 Covariate-adjusted comparisons

Table 6 gives the gender coefficients and standardised differences from the three adjusted logistic models. For conditional willingness, the female-versus-male odds ratio is 1.07 (95% HC1 interval 0.62-1.83), with an adjusted endorsement difference of 1.0 percentage point (-7.4-9.4). The unadjusted and adjusted contrasts differ slightly in sign, but both are small and imprecise. This change is not interpreted as a reversal of a substantive gender relationship.

Table 6. Adjusted gender associations and standardised differences (n = 364)

Outcome	OR	95% HC1 interval	Adjusted gap pp	95% delta interval pp	Holm p
Willingness	1.07	0.62-1.83	1.0	-7.4-9.4	1.000
Trust expectation	1.02	0.67-1.57	0.6	-9.8-11.0	1.000
Discrimination salience	1.00	0.65-1.55	0.1	-10.1-10.3	1.000

Female versus male; all models include age indicators, student status, frequency, any training and rural location. Standardisation averages over the same empirical covariate distribution, held fixed for delta-method intervals. The model gender p-values form a separate three-test Holm family.

For inclusive-design trust expectations, the odds ratio is 1.02 (0.67-1.57), and the adjusted difference is 0.6 points (-9.8-11.0). For discrimination salience, the odds ratio is 1.00 (0.65-1.55), and the adjusted difference is 0.1 points (-10.1-10.3). All Holm-adjusted gender-coefficient p-values are 1.000. Adjustment therefore leaves the substantive interpretation largely unchanged. The measured covariates do not uncover a clear gender contrast that was obscured in the descriptive analysis.

These models do not explain why the gender estimates are small. Similarity could arise from the conditional wording, the composition of the records, shared contextual influences, limited measurement sensitivity or the data-generating process. The available evidence does not discriminate among these explanations. Appendix C reports all coefficients to prevent selective attention to the focal term; contextual coefficients are not promoted into causal explanations or a new family of substantive discoveries.

4.5 Paired willingness-trust-expectation contrast

Within all 385 records, 154 endorse both GLA2 and GLA4, 153 endorse willingness alone, 44 endorse the trust-expectation proposition alone and 34 endorse neither. The paired difference is 109/385, or 28.3 percentage points, with a 95% whole-record bootstrap interval of 21.8-34.8. The exact paired test gives $p < 0.001$. This comparison uses all gender categories and does not reduce the database to the female-male subset.

Figure 2 places the three gender contrasts beside the paired difference. It makes the scale of the observed response-pattern contrast visible without treating the items as a common psychological scale. The gender difference in paired gaps is -1.3 points (95% stratified bootstrap interval -15.1-12.1), leaving substantial uncertainty about whether the within-record contrast varies across the two larger gender categories.

Figure 2. Three gender contrasts and the paired cross-item contrast. Endorsement means scores four or five; intervals are pointwise percentile bootstrap intervals. The paired contrast concerns differently worded statements, not a common-scale trust deficit.

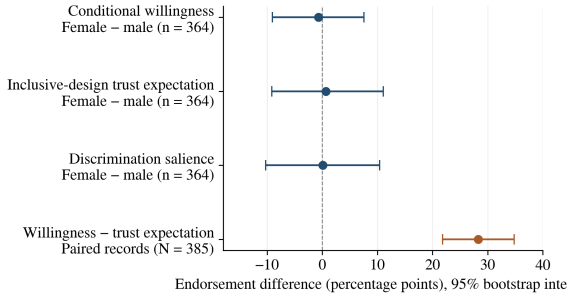


Figure 2. Three gender contrasts and the paired cross-item contrast.

The exploratory gender-by-GLA4 interaction in the willingness model has an odds-ratio ratio of 1.09 (0.60-1.99; raw p =

0.779). This is not evidence of a demonstrated common relationship: the interval accommodates appreciable heterogeneity. The GLA4 coefficient among male records is also imprecise, with an odds ratio of 0.77 (0.52-1.13). Neither estimate supports a claim that inclusive design causes adoption or that gender determines how design affects trust.

4.6 Sensitivity and the limits of equivalence

Table 7 summarises robustness checks. Excluding either never-use response leaves 342 female or male records and yields a willingness gender odds ratio of 0.98 (0.55-1.73). Among current law students, n = 227 within the female-male subset, the corresponding odds ratio is 0.95 (0.48-1.88). These denominators are analytical subsets of the full N = 385; they do not replace it. The similarity of point estimates supports numerical stability across these particular restrictions, while wider student-subset intervals reduce precision.

Table 7. Sensitivity of conditional-willingness gender comparisons

Subset	n	Female versus male OR (95% HC1 interval)
Neither exposure field says never	342	0.98 (0.55-1.73)
Current law students only	227	0.95 (0.48-1.88)

Both rows concern GLA2 endorsement and the female-male subset. The student model omits invariant student status. These subsets do not change the complete descriptive database size of 385.

Table 8. Sensitivity retaining original one-to-five scores (n = 364)

Outcome	Adjusted female-minus-male score difference	95% HC3 interval
Willingness	-0.019	-0.175-0.136
Trust expectation	0.022	-0.150-0.195
Discrimination salience	0.054	-0.111-0.219

Linear models use the primary covariates and HC3 covariance. Score spacing is treated as equal in this sensitivity check. These are separate items, not a composite scale.

Preserving the five response categories through HC3 linear models likewise produces small adjusted gender score differences. This check reduces dependence on the endorsement threshold but assumes equal spacing of score values, so it complements rather than supersedes ordinal distributional comparisons. Neither analysis validates the items as a multi-item scale or addresses unobserved selection.

Pointwise 90% bootstrap intervals for endorsement differences lie inside illustrative bounds of ±10 percentage points for all three focal items, but none lies inside ±5 points. The minimum symmetric half-widths enclosing those intervals are approximately 7.7, 9.5 and 8.8 points. Because these margins were not prospectively justified, this is a precision sensitivity result, not confirmatory evidence of equivalence. Simultaneous equivalence across outcomes, population equality and equality in actual service outcomes have not been established.

5. Discussion

The analysis supports a focused conclusion: the supplied records show high conditional willingness and a substantially lower rate of endorsement for the inclusive-design trust proposition, while female-male contrasts on both statements are small and uncertain. This pattern is more informative when treated as a distinction among measured responses than when converted into a general statement about whether women trust AI less than men. The discussion develops that interpretation while keeping its observational and provenance limits explicit.

5.1 Conditional openness is an incomplete account of adoption

The willingness item supplies two favourable conditions: the service meets a person's needs and appropriate safeguards are present. Respondents can agree with that proposition without endorsing an available service or accepting its actual risks. The result therefore identifies conditional openness within the workbook, not an adoption rate for Bangladesh. It also cannot be used to forecast uptake without measuring service availability, costs, task relevance and the credibility of safeguards.

This distinction has practical value. If decision-makers read the 79.7% figure as readiness for immediate deployment, they may omit the very conditions respondents were asked to assume. A more defensible interpretation treats those conditions as questions for implementation: what qualifies as a relevant service, which safeguards exist, and how can a user verify them? The data do not answer those questions, but the wording makes them central to the meaning of the observed endorsement.

Legal AI trust research also distinguishes reliance from justified reliance. The present evidence cannot show whether a respondent would verify an answer, consult a qualified professional or recognise that a task exceeds the system's capability. The low reliability of the common block reinforces the need to avoid using a general literacy mean as a substitute for these specific behaviours. Future service evaluations should measure decisions and verification in realistic tasks rather than infer them from favourable agreement scores.

5.2 The paired contrast concerns different propositions

The 28.3-point paired contrast is the clearest numerical pattern. More records endorse conditional personal use than endorse the general claim that inclusive design can improve trust. This does not mean willingness exceeds trust by a calibrated amount. GLA4 is about the potential contribution of one design principle; a respondent could be neutral because inclusive design is unfamiliar, because other conditions seem more important or because the statement is too general to evaluate.

The contrast could also reflect differences in item specificity. GLA2 presents a personally relevant scenario with safeguards already assumed. GLA4 presents an abstract proposition without identifying a service or the design changes envisaged. These differences may affect responses independently of any underlying trust relationship. A paired comparison controls for which record supplied the answers, but it cannot remove differences in the questions themselves.

The contribution is therefore methodological as well as substantive. When adjacent items appear under a single heading, readers may assume that all are interchangeable indicators. Here, low internal consistency and a large response-pattern difference both argue for greater specificity. A revised

instrument should separate current institutional trust, trust in technical competence, willingness under named safeguards and actual use. Such separation would permit a stronger test of whether design expectations translate into appropriately calibrated reliance.

5.3 Small gender estimates do not establish structural equality

The descriptive and adjusted contrasts provide little indication of a large observed female-male difference in the focal statements. That finding should constrain interpretations that begin by presuming a female acceptance deficit. A programme justified solely by an assumed lack of women's willingness would find little support in these data. It would still be appropriate to investigate unequal access, exposure to harm and opportunities to challenge a system, because those outcomes remain largely unmeasured.

Equal willingness can coexist with unequal capacity to act. The database records whether someone reports training or frequent AI use, but not whether they can access a device privately, use accessible language, pay for a service or obtain professional assistance. The comparison is also conditional on inclusion in an awareness-eligible database. People without awareness or access are outside that recorded eligibility structure, making extrapolation to all potential legal-service users particularly inappropriate.

The smaller gender categories deserve attention in the interpretation of precision. Their descriptive visibility prevents a binary model from being presented as a complete account of gender. Their sample sizes, however, do not permit stable estimates of distinctive experiences. The correct implication is to improve inclusive sampling and participant-led category design, not to merge them for convenience or claim that a high percentage in eight records demonstrates superior access.

5.4 What adjustment clarifies and what it cannot resolve

Adjustment shows that the focal gender estimates remain small when the specified recorded circumstances are held within a common model. It does not establish that those circumstances are confounders in a causal sense. Training, frequency and professional status may themselves be shaped by unequal opportunities. Conditioning on them can answer a descriptive question about comparable recorded profiles while removing part of the social process that a different research question would seek to understand.

The modest shift in the willingness estimate from -0.7 points unadjusted to 1.0 point adjusted illustrates the importance of reporting both. Neither is a preferred estimate of a causal gender effect. They summarise different descriptive comparisons. Unmeasured institutional affiliation, income, disability, language and legal need could alter either comparison, and unknown recruitment may affect them both. Robust covariance improves the treatment of specified model residuals; it does not repair these design limitations.

Likewise, the exploratory interaction cannot be described as confirmation that inclusive design has an identical role across genders. Its wide interval reflects limited precision for heterogeneity. Improving that precision requires deliberate planning, a clearer design intervention and a sample capable of supporting interaction analysis. Searching alternative thresholds until an interaction becomes significant would provide a weaker basis for inference than acknowledging the

present uncertainty.

5.5 A precision-based interpretation of the null pattern

The sensitivity comparison between five- and ten-point bounds makes the null pattern concrete. The database is more compatible with relatively broad similarity than with a narrow assurance of minimal disparity. A policymaker considering a ten-point difference tolerable may draw a different practical conclusion from one concerned about five points. Statistical analysis cannot choose that ethical threshold retrospectively simply because one bound gives a more convenient result.

For future research, an acceptable margin should be justified through the consequences of a disparity, input from affected people and the purpose of the service. The sample should then be planned to estimate that margin with appropriate precision, including the smaller gender groups. The current intervals can inform such planning, but they cannot become a pre-existing justification for the design that generated them.

Finally, all these interpretations depend on the character of the source data. The codebook's synthetic labels require a more restrained account than a verified field survey would permit. If the numeric records are confirmed as simulated, the study's value is a reproducible analytical demonstration, and all empirical claims about Bangladeshi participants must be removed. If they are documented as collected responses with synthetic identifiers only, recruitment, consent and processing records remain necessary to establish that account. The article reports this uncertainty because it affects the status of every substantive inference.

6. Theoretical and Practical Implications

The article's implications arise from the separation of measured responses and from the limited precision of gender comparisons. They are proposals for theory development and service evaluation, rather than claims that the observed data have demonstrated an effective intervention. The central objective is to connect interest in legal AI with evidence about the conditions that make access meaningful and reliance justified.

6.1 Theoretical implications

First, gender and technology analysis benefits from separating the distribution of willingness from the institutional production of opportunity. A small demographic contrast in conditional intention does not invalidate a contextual account of gendered technology use. It narrows what can be claimed about that particular statement and directs attention to conditions that may intervene between intention and action. This distinction avoids making every unfavourable outcome a personal deficit or every favourable attitude a sign that structural barriers have disappeared.

Second, procedural justice should be measured as more than a favourable attitude toward fairness. An instrument can ask separately whether users can present relevant circumstances, understand the provider's responsibilities, seek human review and obtain a response to a complaint. These variables would allow researchers to test whether particular procedures support legitimate reliance. They would also distinguish approval of a design principle

from confidence in how an actual institution treats a person.

Third, digital inclusion theory can connect access resources to the completion of concrete legal tasks. Future models could measure device privacy, affordability, accessible language, disability accommodation and routes to professional referral, then relate them to observed outcomes. The present training and frequency variables are too broad to serve as complete substitutes. A theory becomes more testable when the instrument records the processes through which its proposed mechanism operates.

6.2 Implications for legal-service providers

Providers should specify safeguards in terms a prospective user can evaluate. A service description can identify the provider, explain the limits of the tool, state how information is handled and show when a qualified human should be consulted. These are implementation proposals, not features verified in the database. They should be tested for comprehension and usefulness, including with people whose access circumstances differ from those of law students and frequent AI users.

Service evaluation should examine both unequal outcomes and unequal opportunities to obtain help. Equal aggregate satisfaction may conceal different error rates, costs or difficulties reaching human review. Algorithmic auditing research provides a basis for documenting responsibilities and checking systems across their lifecycle, while reviewable-decision frameworks emphasise that oversight must enable scrutiny of actual decisions. Applied to legal assistance, this suggests testing the full service journey rather than evaluating an interface alone.[24]

Inclusive design should involve participation in defining needs and evaluating barriers, rather than asking respondents only whether inclusion sounds beneficial. User testing can compare completion rates, error recognition, privacy concerns and the ability to reach a referral. It should report uncertainty and avoid revealing small-group identities. Where performance cannot support a consequential legal task, increasing willingness to use the service would not be an appropriate success criterion.

6.3 Governance and sustainable development

Risk management frameworks and AI ethics recommendations offer relevant principles for organising evaluation, including attention to human oversight and potentially unequal effects. Their use here is normative and methodological; the article does not claim that they are binding domestic law or that compliance has been established. Providers can translate them into documented testing responsibilities, incident handling and periodic reassessment of whom the service reaches and whom it fails.[25]

For SDG-oriented reporting, a useful evaluation should distinguish reach, safe task completion and access to remedy. A willingness percentage belongs to an earlier stage of that sequence. It can help identify interest in a conditional service proposition, but it cannot substitute for evidence that

[24] 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; 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.

[25] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1, NIST 2023); UNESCO, Recommendation on the Ethics of Artificial Intelligence (UNESCO 2022).

legal assistance becomes more accessible or that unequal outcomes decrease. The present study consequently proposes a measurement agenda that connects acceptance research to service consequences without treating attitudinal endorsement as an SDG outcome.

7. Conclusion, Limitations and Future Research

The supplied Group 10 database supports a distinctive analysis of the difference between conditional willingness and inclusive-design trust expectations. Among 385 records, willingness endorsement is 79.7%, while endorsement of the trust proposition is 51.4%. The paired difference is substantial. Female-male point estimates are small across the focal items, but the intervals leave room for differences that may matter in practice. Adjustment and selected sensitivity analyses do not materially change that account.

The conceptual contribution is to keep conditional intention, trust-related beliefs and discrimination salience analytically separate. The methodological contribution is to report uncertainty and sensitivity to practical margins rather than interpret non-significance as equality. Together, these contributions provide a more defensible account of gender comparisons than a composite adoption score or a presumed deficit explanation would offer. They also show why legal-service acceptance research needs information about service conditions and actual outcomes.

Several limitations define the reach of these conclusions. The data are cross-sectional, and the questionnaire does not measure actual adoption, current trust in a specified provider, verified discrimination or completed legal tasks. Single-item outcomes have limited content coverage. The agreement scale may be affected by acquiescence, interpretation and socially desirable responding. No measurement-invariance analysis is possible for a validated multi-item construct because no such construct is established here.

Recruitment, clustering, contact procedures and field dates are undocumented. The sample contains many law students and awareness-eligible records, limiting generalisation to people facing legal problems or lacking digital access. The female-male regression excludes 21 other or non-disclosing gender responses from that particular comparison; their experiences cannot be inferred from the two larger categories. Covariate adjustment is incomplete, and some measured covariates may be consequences of social opportunity rather than external confounders.

Experience fields are discordant, and the sensitivity exclusions cannot identify the correct original response. Bootstrap intervals and sandwich covariance rely on an independent-record analytical model; they do not account for unknown recruitment clusters. Multiple-testing adjustment covers specified three-item families, not every exploratory calculation or interpretive comparison. Illustrative equivalence margins were not established before analysis, and their sensitivity results must not be presented as a confirmatory finding of parity.

Most fundamentally, the codebook labels identifiers and comments synthetic while leaving numeric-response provenance insufficiently documented. This is a reporting boundary rather than a statistical detail. A final empirical submission requires confirmation of whether the responses represent human participants, a generated teaching database or a mixture. If simulated, the manuscript must be submitted only in a form that explicitly identifies its methodological-demonstration status and

does not represent the results as participant evidence.

Future research should recruit transparently across relevant service users, include people with limited awareness and design gender categories with participant input. It should measure specific trustees, test comprehension of safeguards and observe task completion, verification, referrals and review. A prospective protocol should justify disparity margins, power interaction analyses and protect smaller groups. Such a design could determine whether conditional openness becomes safe and equitable legal access—an outcome the present database cannot establish.

8. Declarations

Ethical approval and oversight

Formal ethical approval was not obtained because, at the time of the study, the authors' institutional setting did not provide an accessible formal review mechanism applicable to this non-interventional adult questionnaire research. The study nevertheless followed participant-protection safeguards, including informed consent, voluntary participation, the opportunity to discontinue, an anonymous questionnaire and aggregate reporting. The methodological safeguards are detailed in Section 3.8. The absence of an accessible review mechanism describes the institutional context; responsibility for ethical conduct remained with the researchers.

Informed consent

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

Data availability

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

Author contributions

Md. Didarul Islam: 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

Ajzen I, ‘The Theory of Planned Behavior’ (1991) 50 Organizational Behavior and Human Decision Processes 179

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

Aoki N, ‘An Experimental Study of Public Trust in AI Chatbots in the Public Sector’ (2020) 37 Government Information

Quarterly 101490

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

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

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

Crenshaw K, 'Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics' [1989] University of Chicago Legal Forum 139

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

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

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

Faulkner W, 'The Technology Question in Feminism: A View from Feminist Technology Studies' (2001) 24 Women's Studies International Forum 79

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

Helsper EJ, 'A Corresponding Fields Model for the Links Between Social and Digital Exclusion' (2012) 22 Communication Theory 403

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

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

Lakens D, 'Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses' (2017) 8 Social Psychological and Personality Science 355

Long D and Magerko B, 'What Is AI Literacy? Competencies and Design Considerations' in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (ACM 2020)

MacKinnon JG and White H, 'Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties' (1985) 29 Journal of Econometrics 305

Mann HB and Whitney DR, 'On a Test of Whether One of Two Random Variables Is Stochastically Larger than the Other' (1947) 18 Annals of Mathematical Statistics 50

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

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

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

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

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

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

Tyler TR, *Why People Obey the Law* (Yale University Press 1990)

UN General Assembly, 'Transforming Our World: The 2030 Agenda for Sustainable Development' (25 September 2015) UN Doc A/RES/70/1

UNESCO, 'AI for All: Shaping Bangladesh's Future with Ethical AI' (22 May 2025) <https://www.unesco.org/en/articles/ai-all-shaping-bangladeshs-future-ethical-ai> accessed 25 October 2026

——, *Recommendation on the Ethics of Artificial Intelligence* (UNESCO 2022)

Venkatesh V and Morris MG, 'Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior' (2000) 24 MIS Quarterly 115

Venkatesh V and others, 'User Acceptance of Information Technology: Toward a Unified View' (2003) 27 MIS Quarterly 425

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

Appendix A. Questionnaire wording and analytical coverage

The following statements reproduce the supplied focal questionnaire wording. Their analytical roles follow their content rather than the group heading. The original one-to-five agreement coding is retained throughout, with endorsement thresholds used only where expressly stated.

Table A1. Exact focal questionnaire statements

Code	Statement
GLA1	I believe people of different genders should have equal access to AI-enabled legal services.
GLA2	I would use an AI legal service if it met my needs and appropriate safeguards were in place.
GLA3	Concerns about discrimination affect my willingness to use AI legal services.
GLA4	Inclusive design can improve trust in AI legal services.
GLA5	AI legal services should be tested for unequal outcomes across relevant groups.

Response scale: one strongly disagree, two disagree, three neutral, four agree, five strongly agree.

The common questionnaire includes GBA and WLE blocks relevant to the other Group 10 topics. Their reliability is reported for audit purposes, but they are not transformed into the primary outcomes or explanatory mechanisms of this article. Optional contextual comments are excluded because the codebook identifies them as synthetic.

Appendix B. Reproducibility specification

Read the workbook's Data sheet without dropping rows. Check unique Respondent_ID, the three eligibility fields and the one-to-five range for AIL1-5, GLA1-5, GBA1-5 and WLE1-5. Tabulate all four gender responses. Define the female-male subset using only exact Female and Male values. Define endorsement as an original score greater than or equal to four. Preserve neutral responses in full-distribution and rank analyses.

For stratified resampling, draw 169 female and 195 male records with replacement independently in each replicate. Compute the female-minus-male endorsement difference and use the 2.5th and 97.5th percentiles for the 95% interval. Use the fifth and 95th percentiles for the 90% precision sensitivity interval. For paired whole-database contrasts, resample 385 complete records and compute the mean of the within-record GLA2-minus-GLA4 binary differences. Use 10,000 resamples and NumPy default_rng with seed 20261025 in the sequence described in the analysis script.

The logistic design matrix contains an intercept, female indicator, four age indicators with 18-20 as reference, current law-student indicator, frequency coded zero never through four daily, any-training indicator and upazila-rural indicator. Maximise the Bernoulli log likelihood using BFGS with an analytical gradient. Compute HC1 covariance from record score outer products and the inverse Hessian, multiplied by n/(n-k). For standardisation, average probabilities after replacing the gender column with zero and then one; use the delta method with the covariate distribution fixed.

The exploratory interaction model additionally includes GLA4 minus three and its product with the female indicator. Its gender coefficient refers to GLA4 = 3, and the interaction is an odds-ratio ratio per score point. Do not interpret its coefficient as a treatment effect. In the law-student sensitivity model remove the invariant status indicator. HC3 score regressions use the same primary covariates and leverage-adjusted residual covariance. Software convergence checks for the three primary logistic models yield maximum score components below 0.000001; independent coefficient checks differ by less than 0.000001.

All intervals are pointwise. Holm adjustment is applied separately to the three rank-test p-values and the three primary model gender-coefficient p-values. Sensitivity and interaction p-values are descriptive. The provenance caveat must accompany any reused table or result; numerical reproducibility cannot establish that records originated from human participants.

Appendix C. Complete model estimates

Table C1 provides the contextual coefficients alongside the focal gender term. Reference categories and coding are listed in Appendix B. These coefficients are included for transparency and are not treated as a separate set of tested causal mechanisms.

Table C1. Complete adjusted model coefficients (n = 364)

Predictor	GLA2 OR (95% CI)	GLA4 OR (95% CI)	GLA3 OR (95% CI)
Female versus male	1.07 (0.62-1.83)	1.02 (0.67-1.57)	1.00 (0.65-1.55)
Age 21-23 versus 18-20	0.78 (0.35-1.70)	1.04 (0.57-1.91)	1.22 (0.66-2.23)
Age 24-26 versus 18-20	1.12 (0.43-2.90)	0.86 (0.43-1.72)	1.96 (0.96-3.96)
Age 27-30 versus 18-20	0.27 (0.09-0.75)	0.74 (0.30-1.81)	1.13 (0.48-2.71)
Age 31+ versus 18-20	0.76 (0.24-2.40)	0.73 (0.29-1.84)	1.50 (0.60-3.77)
Current law student	0.71 (0.40-1.24)	0.68 (0.44-1.05)	1.33 (0.85-2.08)
AI use frequency step	1.04 (0.79-1.37)	1.17 (0.95-1.45)	1.40 (1.13-1.73)
Any formal AI training	1.13 (0.65-1.96)	1.03 (0.66-1.59)	1.49 (0.96-2.34)
Upazila-rural location	1.80 (0.83-3.91)	0.91 (0.52-1.57)	1.51 (0.86-2.66)

All intervals use HC1 covariance and are pointwise. Intercepts are omitted. Frequency is per one-category increase; training contrasts any with none, status contrasts current law students with others, and location contrasts upazila-rural with all other locations.

Covariate coefficients describe the fitted associations conditional on this specification. Intercepts are omitted from the table because their reference profile is not a substantive population benchmark. The models contain no survey weights, random effects or cluster corrections because the required design information is unavailable.