Research Square

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

Trust Conditions and Conditional Willingness to Use AI Legal Information in Bangladesh

Dr Kazi Abdul Mannan

drkaziabdulmannan@gmail.com

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

Chowdhury Niamul Fahmid

Shanto-Mariam University of Creative Technology https://orcid.org/0009-0009-3939-0564

Research Article

Keywords: artificial intelligence, legal information, trust, verification, conditional willingness, access to justice, Bangladesh

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11685610/v1

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

Additional Declarations: The authors declare no competing interests.

Trust Conditions and Conditional Willingness to Use AI Legal Information in Bangladesh

Chowdhury Niamul Fahmid

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

Email: arikniamul@gmail.com ORCID: 0009-0009-3939-0564

Prof. Dr Kazi Abdul Mannan

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

Corresponding author: drkaziabdulmannan@gmail.com ORCID: 0000-0002-7123-132X

Abstract

A person can express willingness to use an artificial intelligence (AI) legal service under transparent conditions while remaining cautious about acting on its information. This cross-sectional study distinguishes conditional willingness, trust expectations and verification caution using 385 adult questionnaire records from Bangladesh. We report full response distributions, paired comparisons and exploratory adjusted models of willingness attributed to transparent explanations. The willingness statement was endorsed by 337 respondents (87.5%); trust conditional on clear sources was endorsed by 260 (67.5%) and trust conditional on qualified professional oversight by 266 (69.1%). Within respondents, willingness endorsement exceeded these two trust conditions by 20.0 and 18.4 percentage points, respectively (both Holm-adjusted $p < 0.001$). Of the 337 respondents who endorsed conditional willingness, 233 also endorsed the professional oversight condition. In the adjusted original-rating model, perceived access utility ($b = 0.275$, 95% CI 0.111 to 0.439) and verification caution ($b = 0.154$, 95% CI 0.060 to 0.248) were positively associated with conditional willingness (both adjusted $p = 0.003$). Neither association met the adjusted threshold when willingness was dichotomised. The findings distinguish openness to a transparent service from confidence in an answer or readiness to act on it. They motivate separate evaluation of source checking, professional escalation and service use. The study does not identify the causal effect of transparency or estimate actual adoption or national prevalence.

Keywords: artificial intelligence; legal information; trust; verification; conditional willingness; access to justice; Bangladesh.

1. Introduction

A person trying to find legal assistance may first need to identify the relevant issue, learn which organisation can help, or find a document they can read. An AI service may reduce the effort of taking that first step. The same service may give an answer that sounds definite when an important fact is missing or a cited rule does not apply. Access to information and the reliability of information are therefore distinct properties of a service. Expanding access to an answer is useful only to the extent that people can determine what the answer is suitable for and what should be checked before they act. Evidence from evaluations of legal research tools illustrates that incorrect or unsupported answers remain a practical concern even when a system is designed for legal work.[1]

A person can be willing to try a tool without trusting it to resolve a consequential problem. Someone might use a service to find terminology, then check a primary source or consult a professional. Another might trust the tool to provide office hours but not to identify a filing deadline. A survey that equates willingness with confidence in the accuracy of a particular answer risks missing these differences. The distinction is particularly important when a question is conditional. Agreement that transparent explanations increase willingness does not identify willingness in the absence of those explanations, nor does it document any actual decision to use a service.

The Group 2 questionnaire distinguishes trust under identifiable sources, trust under qualified oversight, expectations of clear limits, caution about acting without verification, and willingness under transparent explanations. These judgements can occur together, but their frequency and covariation are empirical questions. A total score would mix statements about what institutions should provide with statements about the respondent's own willingness or caution. The present article instead makes the differences in item content the starting point for analysis.

We ask how often each condition is endorsed; how endorsement of willingness compares within respondents with endorsement of source identification, professional oversight, limits and verification caution; and whether access utility, trust conditions or caution covary with the willingness rating. The joint analysis further shows how many people endorse professional oversight and conditional willingness together. This makes it possible to distinguish the prevalence of both positions from any association between them.

The study concerns the AI-based legal information topic assigned to Chowdhury Niamul Fahmid in the Group 2 workbook. The common instrument also supports separate articles on perceived usefulness and legal-aid accessibility. Shared records do not provide independent replication. Our focus is the relation between conditional willingness and

[1] Varun Magesh and others, ‘Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools’ (Stanford Law School) <https://law.stanford.edu/publications/hallucination-free-assessing-the-reliability-of-leading-ai-legal-research-tools/>.

safeguards, rather than measuring changes in access to justice or evaluating the accuracy of a particular service.

2. Literature Review and Theoretical Foundation

The framework distinguishes perceived utility, conditions of trust and appropriate reliance. Utility concerns what a service can help a person accomplish. Trust conditions concern information and oversight that respondents say would matter to them. Appropriate reliance concerns the action a person takes given the system's capabilities and the consequences of error. The questionnaire captures stated judgements about these concepts; it does not record reliance behaviour.

2.1 Access utility

Perceived usefulness concerns an expected contribution to task performance. In the present instrument, the relevant tasks involve access to basic information, speed, convenience and the potential to help people facing barriers. These expectations can motivate willingness even when a person believes that an answer needs checking. For example, an initial response may be useful because it identifies an organisation to contact. Its value in that task does not show that the same response can answer a case-specific legal question.[2]

Access to justice encompasses more than obtaining preliminary information. A user may need to understand a procedure, gather documents or obtain qualified assistance. The questionnaire does not measure whether those steps were completed. In the model, perceived access utility is the equal-weight mean of four statements and describes broad expectations about preliminary information. The separate item about considering use is excluded so the utility predictor does not directly include a willingness judgement.

2.2 Trust conditions and verification

Trust in automation is relevant to whether people rely on a system in situations where they cannot fully evaluate its operation. The design goal is appropriate reliance: use where the system can support the task and caution where its limitations matter. A clearly identified source can help a user check an answer, but does not guarantee that the source is current or correctly applied. A professional oversight arrangement can help where it gives a qualified person access to evidence and authority to correct a conclusion. Merely naming a reviewer or adding a citation does not establish that those processes function.[3]

---

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

[3] John D Lee and Katrina A See, 'Trust in Automation: Designing for Appropriate Reliance' (2004) 46 Human Factors 50; Magesh and others (n 1).

Verification caution records reluctance to act on information without checking it. Such caution can be compatible with trying a service: a user may view it as a starting point for inquiry rather than the basis for action. A positive association with willingness would be consistent with that interpretation, but could also reflect a common tendency to endorse safeguards. A weak association with the trust-condition mean would not show that source identification and oversight are dispensable. The data record attitudes to broad conditions rather than responses to changes in an actual service.

2.3 Analytical expectations

Perceived utility may track willingness because the value of preliminary information can make a service worth considering. Verification caution may track willingness when respondents imagine a workflow that allows them to check the result. It could also be associated with lower willingness if checking appears costly or difficult. The two trust items concern source and oversight conditions, while the outcome concerns explanations. These may be related, but a person can endorse one condition without endorsing another.

We examine these possibilities using observed responses. The outcome asks whether transparent explanations increase willingness. It does not measure baseline trust, current adoption, the probability of future use or change from a measured baseline. The explanatory mean does not measure trust in a named service. We use the terms conditional trust and conditional willingness only to preserve the conditions expressed by the respondents.

3. Methodology

The analysis uses a cross-sectional questionnaire and reports the achieved sample, observed measures and limits of inference. Within-person comparisons preserve the pairing of responses. The reporting identifies the available recruitment information and distinguishes statistical uncertainty from representativeness.[4]

3.1 Source and screening

The Group 2 workbook contains a Data sheet, README and codebook for the common Q2 questionnaire. The unit is an adult respondent. All 385 records contain affirmative consent, adult eligibility and AI awareness. We checked sequence uniqueness, valid values for all 20 agreement items and completeness of numerical and profile fields. No duplicate response and profile patterns were found after excluding the sequence and contextual text. No record was removed. “None” remained a valid tool response.

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

The study team confirmed that the numerical responses were empirical and identified the synthetic descriptions of the sequence and contextual fields as metadata labelling errors. Contextual text was not analysed. The recruitment frame, invitation denominator, exact dates and mode of distribution were not documented in the available materials. We report no response rate, sampling weight or nationally representative margin of error.

3.2 Measures

Table 1 gives the analytical roles of the items. The access-utility mean uses PUA1-4. TAIL1 and TAIL2 form the observed trust-conditions mean; they ask about identified sources and professional oversight. TAIL4 enters as a separate verification-caution rating, and TAIL5 is the outcome. The limitations statement, TAIL3, is reported descriptively and is included in paired comparisons. All ratings retain their original 1-5 direction.

Table 1. Analytical roles of questionnaire items

Measure	Items	Interpretation
Access utility	PUA1-4	Mean of accessibility, speed, barrier reduction and convenience
Trust conditions	TAIL1-2	Mean of source and professional oversight conditions
Verification	TAIL4	Caution about acting without verification
Conditional willingness	TAIL5	Willingness attributed to transparent explanations
Adjustment	AIL1-3 and profile	General AI orientation and categorical background terms

TAIL5 does not enter the predictors.

AIL1-3 provide a general AI orientation summary covering reported recognition of inaccuracy and bias and reported verification. This is not a skills test. Appendix A provides exact wording. We report the TAIL block's coefficient alpha and rank correlations as diagnostics. A low alpha does not establish separate factors or justify a formative score. The substantive differences among the statements warrant separate reporting; no structural or mediation model is estimated.[5]

3.3 Statistical analysis

We define endorsement as agree or strongly agree and report all original categories. Neutrality is included in non-endorsement, which is not interpreted as opposition. We cross-tabulate TAIL2 and TAIL5. The four paired contrasts compare TAIL1-4 with TAIL5. The estimated difference is the comparator-only count minus the willingness-only count, divided by 385. Exact McNemar tests use discordant pairs. Pointwise percentile intervals

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

use 10,000 whole-record resamples with replacement and seed 20261104. Holm adjustment covers all four contrasts.[6]

The exploratory linear model uses the original TAIL5 rating. The focal terms are access utility, trust conditions and verification caution. The adjusted specification adds general AI orientation and categorical terms for status, location, training and AI-use frequency. Coefficients are unstandardised, with HC3 standard errors and pointwise normal-approximation intervals. Holm correction applies to the three focal coefficients within each model. We repeat the adjusted specification with a binary endorsement outcome and among the law students. The first sensitivity model describes probability differences; the second uses an overlapping subset and is not a replication or test of subgroup differences.[7]

3.4 Ethics and data protection

Ethical approval for this study was obtained from the Center for Academic & Professional Career Development and Research (CAPCDR), an external institutional body, under approval reference CAPCDR: SMUCT/LLB/SS-0020260908-17. External ethical review was sought because Shanto-Mariam University of Creative Technology does not have an institutional ethics committee responsible for reviewing this type of research.

4. Results

The results describe sample composition, item distributions, within-person overlap and adjusted associations. All primary analyses use 385 records.

4.1 Sample profile

Table 2 shows age, status, location and training. The largest age group was 21-23 (169, 43.9%). There were 227 law students, 73 law graduates, 58 advocates or legal professionals and 27 other respondents. Dhaka accounted for 150 records. No formal AI training was reported by 205, short orientation by 143 and course or certified training by 37.

Table 2. Respondent profile

Characteristic	Category	n (%)
Age	21-23	169 (43.9%)
Age	24-26	84 (21.8%)
Age	18-20	53 (13.8%)

[6] Quinn McNemar, 'Note on the Sampling Error of the Difference between Correlated Proportions or Percentages' (1947) 12 Psychometrika 153; Sture Holm, 'A Simple Sequentially Rejective Multiple Test Procedure' (1979) 6 Scandinavian Journal of Statistics 65.

[7] James G MacKinnon and Halbert White, 'Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties' (1985) 29 Journal of Econometrics 305; Holm (n 6).

Characteristic	Category	n (%)
Age	31+	46 (11.9%)
Age	27-30	33 (8.6%)
Status	Law student	227 (59.0%)
Status	Law graduate	73 (19.0%)
Status	Advocate or legal professional	58 (15.1%)
Status	Other relevant respondent	27 (7.0%)
Location	Dhaka	150 (39.0%)
Location	District town	86 (22.3%)
Location	Other city corporation	78 (20.3%)
Location	Upazila-rural area	61 (15.8%)
Location	Other	10 (2.6%)
Training	No	205 (53.2%)
Training	Yes, short orientation	143 (37.1%)
Training	Yes, course or certified training	37 (9.6%)

The sample contains a large legal education and practice component. It does not identify a representative cohort of justice seekers, and the questionnaire does not verify use of a named legal service.

4.2 Item distributions

Table 3 reports all categories. Source-conditional trust was endorsed by 260 (67.5%) and oversight-conditional trust by 266 (69.1%). Clear limitations were endorsed by 325 (84.4%) and verification caution by 329 (85.5%). Conditional willingness was endorsed by 337 (87.5%); 45 chose neutrality and three disagreed.

Table 3. Distribution of trust conditions and conditional willingness ratings

Item	1	2	3	4	5	Endorsed
TAIL1	0	14	111	173	87	260 (67.5%)
TAIL2	0	24	95	182	84	266 (69.1%)
TAIL3	0	5	55	172	153	325 (84.4%)
TAIL4	0	3	53	163	166	329 (85.5%)
TAIL5	0	3	45	176	161	337 (87.5%)

Endorsement combines categories 4 and 5.

The TAIL block alpha was 0.192. Rank correlations between willingness and the other items ranged from −0.023 for oversight to 0.194 for verification caution. Weak covariation and distinct item content argue against interpreting the overall mean as a validated trust score.

4.3 Joint endorsement and paired contrasts

Table 4 shows that 233 respondents (60.5%) endorsed both the professional oversight condition and conditional willingness. Another 104 endorsed willingness without endorsing oversight, 33 showed the reverse pattern and 15 endorsed neither. Among oversight endorsers, 87.6% endorsed willingness, compared with 87.4% of the remaining respondents. The similarity shows why substantial overlap need not imply an association.

Table 4. Joint endorsement of professional oversight and conditional willingness

Oversight condition	Willingness not endorsed	Willingness endorsed	Total
Not endorsed	15 (3.9%)	104 (27.0%)	119 (30.9%)
Endorsed	33 (8.6%)	233 (60.5%)	266 (69.1%)
Total	48 (12.5%)	337 (87.5%)	385 (100.0%)

Percentages are of all 385 records.

Table 5 compares the four statements with willingness. Source-conditional trust was lower by 20.0 percentage points (95% CI 14.3 to 25.7 in the willingness-minus-source direction); oversight-conditional trust was lower by 18.4 points (95% CI 12.7 to 24.2). Both adjusted p values were below 0.001.

Table 5. Paired contrasts with conditional willingness

Contrast	Comparator only	Willingness only	Gap (pp)	95% CI (pp)	Holm p
TAIL1 − TAIL5	30	107	-20.0	-25.7 to -14.3	<0.001
TAIL2 − TAIL5	33	104	-18.4	-24.2 to -12.7	<0.001
TAIL3 − TAIL5	35	47	-3.1	-7.8 to 1.3	0.448
TAIL4 − TAIL5	37	45	-2.1	-6.8 to 2.3	0.448

Pointwise percentile intervals and Holm-adjusted exact McNemar tests.

Clear limitations and verification caution differed from willingness by −3.1 and −2.1 points, respectively, in the comparator-minus-willingness direction. Their intervals included zero, and both adjusted p values were 0.448. This does not establish equivalence. One hundred and twenty respondents (31.2%) endorsed all five statements; that count describes a response pattern, not a standard for safe use.

4.4 Adjusted models

Table 6 reports the estimates. Access utility was associated with the original willingness rating (b = 0.275, 95% CI 0.111 to 0.439), as was verification caution (b = 0.154, 95% CI 0.060 to 0.248). Both Holm-adjusted p values were 0.003. The trust-conditions estimate was 0.005 (95% CI −0.114 to 0.124; adjusted p = 0.936). The adjusted model explained 14.9% of observed variance.

Table 6. Associations with the conditional willingness response

Model (n)	Term	Estimate	95% CI	Holm p	R^2
Unadjusted (385)	Access utility	0.346	0.190 to 0.502	<0.001	0.083
Unadjusted (385)	Trust conditions	0.046	-0.067 to 0.159	0.426	0.083
Unadjusted (385)	Verification	0.178	0.087 to 0.269	<0.001	0.083
Adjusted (385)	Access utility	0.275	0.111 to 0.439	0.003	0.149
Adjusted (385)	Trust conditions	0.005	-0.114 to 0.124	0.936	0.149
Adjusted (385)	Verification	0.154	0.060 to 0.248	0.003	0.149
Binary endorsement (385)	Access utility	0.079	0.003 to 0.156	0.082	0.097
Binary endorsement (385)	Trust conditions	0.015	-0.044 to 0.073	0.617	0.097
Binary endorsement (385)	Verification	0.059	0.009 to 0.109	0.061	0.097
Law students (227)	Access utility	0.223	-0.005 to 0.451	0.110	0.195
Law students (227)	Trust conditions	0.013	-0.145 to 0.171	0.871	0.195
Law students (227)	Verification	0.172	0.046 to 0.299	0.023	0.195

HC3 intervals are pointwise. Binary endorsement coefficients are probability differences; other coefficients are on the 1-5 scale.

In the binary endorsement model, the access-utility and verification estimates remained positive, but neither met the adjusted threshold ($p = 0.082$ and 0.061). In law students, verification caution retained an association with the original rating ($b = 0.172$, 95% CI 0.046 to 0.299; adjusted $p = 0.023$), while utility did not. These analyses do not establish that the effects differ by respondent status. They show that inference depends in part on whether distinctions among the original response levels are retained.

5. Discussion

The key finding is that willingness under transparent explanations was more widely endorsed than increased trust under either source identification or qualified oversight. The within-person comparisons show that the difference is not merely a comparison of different groups. The conditions and outcomes are nevertheless worded differently. The study does not show that explanations are more effective than oversight, or that users would choose one safeguard over another.

5.1 Willingness and appropriate reliance

The high endorsement of verification caution alongside willingness suggests that openness should not be interpreted as readiness to act on an unverified answer. A respondent can be interested in preliminary assistance and insist on checking anything consequential. The positive original-rating association between caution and willingness is consistent with that interpretation. It is also compatible with other explanations, including shared response tendencies or an unmeasured preference for services with safeguards. No observed

behaviour distinguishes these explanations. An evaluation would need to record what a user actually checks and which decisions they defer to a qualified person.[8]

The binary model illustrates a measurement issue. When nearly nine in ten respondents endorse the outcome, collapsing agreement and strong agreement discards an important source of variation. The reduced statistical evidence does not demonstrate that the original-scale model was wrong; the two specifications ask different questions. It does, however, limit claims that the findings are stable across operational definitions of willingness. The adjusted models should be read alongside the full response distributions.

5.2 Oversight and source identification

The joint results show that many respondents endorse oversight and willingness together. Nearly identical willingness rates among oversight endorsers and non-endorsers show that overlap alone does not establish covariation. The adjusted trust-conditions estimate was also close to zero. This is not evidence against providing oversight: the study measures agreement with a statement about oversight, not exposure to a service that offers it. An institution should assess oversight according to whether the reviewer can detect and correct errors, intervene in a timely way and explain the decision. Whether users find oversight reassuring is a separate outcome.

Source identification works similarly. A link can help someone check an answer, but that requires understanding which proposition it supports and whether the authority is applicable. A system can supply an authentic source while misstating its meaning. A useful verification interface should distinguish source retrieval from a conclusion and indicate which assumptions affect applicability.[9]

5.3 Utility and the access question

The original-scale utility association is consistent with the possibility that respondents who see value in preliminary information are more open to transparent services. It does not demonstrate that a deployed service reduces a barrier. Better access can depend on language, affordability, connectivity and whether a person can obtain further assistance when the first answer is insufficient. The study does not measure those outcomes. A legal-service evaluation should establish whether a person locates appropriate assistance or completes a defined step instead of treating willingness as evidence of access to justice.[10]

[8] Lee and See (n 3).

[9] Magesh and others (n 1).

[10] Davis (n 2).

6. Theoretical and Practical Implications

The findings support measuring intended use, trust in specific functions and verification behaviour separately. These measures answer different questions about a service. A person may use a tool, distrust one answer and make a sound choice by consulting an appropriate source. Evaluating only use or a general confidence score can miss that successful pattern of cautious reliance.

6.1 Service design and assessment

A service could state what it is designed to do, provide the source and date for material claims, identify uncertain assumptions and offer a route to human support. These features should be tested using tasks that require a user to identify when guidance is preliminary, verify a proposition and recognise when a qualified person is needed. Evaluation can also record whether the escalation route is usable by someone with limited time or access to digital services. Transparency is useful when it helps a person take the next appropriate step, even if it does not increase confidence.[11]

6.2 Policy and research

A trial could independently vary source visibility, explanation format and availability of qualified review while holding the question and underlying answer constant. Outcomes could include comprehension, correct verification, willingness to try the service and actual selection of a next step. This would distinguish assurance that raises confidence from a feature that improves decision quality. The results could then inform service design without assuming that the most popular feature is the most effective safeguard.

7. Conclusion Limitations and Future Research

Among 385 adults in the Group 2 questionnaire, conditional willingness was more common than trust conditional on clear sources or professional oversight. Perceived utility and verification caution were associated with the original willingness rating, but neither met the adjusted threshold for binary endorsement. The evidence supports treating openness to preliminary assistance separately from confidence in the result and readiness to act.

The sample is concentrated in legal education and practice. Recruitment mode and dates and an invitation denominator were not established in the materials. The questionnaire does not verify service use, identify a named service or measure a change after a transparency intervention. It contains heterogeneous single items and observed summaries, and the models cannot establish direction or address all sources of confounding. Future studies

[11] Lee and See (n 3); Magesh and others (n 1).

should recruit users with documented legal information needs, vary safeguards in specific scenarios and measure accuracy of verification and subsequent behaviour.

Declarations

Ethical approval and oversight

Ethical approval for this study was obtained from the Center for Academic & Professional Career Development and Research (CAPCDR), an external institutional body, under approval reference CAPCDR: SMUCT/LLB/SS-0020260908-17. External ethical review was sought because Shanto-Mariam University of Creative Technology does not have an institutional ethics committee responsible for reviewing this type of research.

Informed consent

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 funding or competing interest was disclosed in the available study materials.

Author contributions

Chowdhury Niamul Fahmid is the lead student author. Prof. Dr Kazi Abdul Mannan is the supervising co-author and corresponding author. The authors are responsible for review and approval of the final manuscript.

Generative AI disclosure

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

Bibliography

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

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

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

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

Magesh V and others, ‘Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools’ (Stanford Law School) <https://law.stanford.edu/publications/hallucination-free-assessing-the-reliability-of-leading-ai-legal-research-tools/>

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

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’ (2007) 4 PLoS Medicine e296

Appendix A Questionnaire wording

Table A1 gives the exact wording of the focal statements and adjustment items.

Table A1. Exact questionnaire wording

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.
PUA1	AI-based legal information can make basic legal guidance easier to access.
PUA2	AI legal assistants can reduce the time needed to obtain preliminary legal information.
PUA3	AI-enabled services can be useful for people who face barriers to conventional legal services.
PUA4	AI legal tools can improve the convenience of obtaining basic legal information.
TAIL1	I would trust AI legal information more when its sources are clearly identified.
TAIL2	I would trust an AI legal service more if a qualified legal professional provides oversight.
TAIL3	I believe AI legal tools should clearly state their limitations.
TAIL4	I am cautious about acting on AI-generated legal information without verification.
TAIL5	Transparent explanations increase my willingness to use AI legal services.

Source: Group 2 Codebook. Ratings range from 1 strongly disagree to 5 strongly agree.

TAIL5 is the outcome and does not enter any predictor.