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.

Perceived Authorship Responsibility and Reported Practices in AI Assisted Legal Writing 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

Mst. Sadia Rahman Arpita

Shanto-Mariam University of Creative Technology https://orcid.org/0009-0004-6111-3495

Research Article

Keywords: authorship responsibility, legal education, generative artificial intelligence, academic integrity, disclosure, legal verification, Bangladesh

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11554213/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.

Perceived Authorship Responsibility and Reported Practices in AI Assisted Legal Writing in Bangladesh

Mst. Sadia Rahman Arpita
Department of Law, Shanto-Mariam University of Creative Technology, Dhaka, Bangladesh
Email: sadiarahmanarpita@gmail.com
ORCID: https://orcid.org/0009-0004-6111-3495
Prof. Dr Kazi Abdul Mannan
Adjunct Professor, Department of Business Administration, Shanto-Mariam University of Creative Technology, Dhaka, Bangladesh
Email: drkaziabdulmannan@gmail.com
ORCID: https://orcid.org/0000-0002-7123-132X

Abstract

Generative artificial intelligence can assist legal writing while leaving unresolved how writers understand responsibility for accuracy, verification, disclosure and explanation. These responsibilities should not be assumed to form a single construct or to correspond directly with reported practice. This study analyses 385 adult, AI-aware survey records from Bangladesh, including 245 law students, using separate authorship statements and content-matched reports of verification and disclosure. Analyses combine response distributions, paired endorsement contrasts, joint patterns, ordinal correlations and an exploratory proportional-odds model of perceived disclosure responsibility. Responsibility for accuracy was endorsed by 89.6%, the duty to verify authorities by 85.2%, required disclosure by 66.8%, explaining and defending submitted text by 67.8%, and avoiding reliance on ununderstood writing by 60.3%. Verification duty and reported verification had similar endorsement proportions, yet 95 records differed in binary endorsement. Required disclosure and reported disclosure differed in 170 records. These response combinations do not establish misconduct or an intention–behaviour gap. Seven selected ordinal associations were small and uncertain after Holm adjustment. In the adjusted disclosure-responsibility model, policy awareness, reported disclosure and accuracy responsibility had odds ratios of 1.20, 1.16 and 1.22 respectively, with intervals spanning one. Student-restricted estimates were similarly uncertain. The findings distinguish widely endorsed principles from their cross-item alignment rather than validate a causal policy-to-integrity mechanism. They support evaluating task-specific disclosure, source verification and explanation through observable assessment activities. Undocumented recruitment and fieldwork dates, mixed respondent status, single-item measures and source-documentation limitations constrain inference. The article provides a measurement-based foundation for legal-writing education without equating favourable self-reports with demonstrated compliance.

Keywords: authorship responsibility; legal education; generative artificial intelligence; academic integrity; disclosure; legal verification; Bangladesh

1. Introduction

AI-assisted legal writing raises several related questions about human responsibility. Is the writer accountable for factual accuracy? Must a cited authority be checked? When should assistance be acknowledged? Can the writer explain an argument that a tool helped produce? Agreement with one principle does not necessarily establish how a person understands the others. Nor does agreement demonstrate that a particular text has been verified or that a disclosure meets an applicable rule. These distinctions matter when legal educators interpret surveys and design assessments.

The concern about accuracy has an empirical basis. Dahl and colleagues examined legal hallucinations in large language models using verifiable legal questions.[1] Their work identifies a problem in particular models and tasks, rather than a universal error rate for every current system. It nevertheless illustrates why a plausible legal statement should be assessed against authoritative sources. Comparative professional guidance likewise addresses competence, confidentiality and other responsibilities when lawyers use generative AI.[2] Such guidance informs discussion here but is not presented as binding law governing every Bangladesh student assignment.

Academic writing also serves a learning function. An accurate final answer may not show whether a student can identify an issue, locate an authority or defend an interpretation. Disclosure can help an assessor understand how the work was produced, while explanation can help assess what the student understands. Both require clarity about the particular task and the assistance permitted. A general statement that a person remains responsible therefore needs to be connected to concrete expectations rather than treated as a complete account of responsible writing.

The empirical problem is how perceived authorship responsibility is distributed across its component statements and how those statements relate to reported practice. A survey may find widespread agreement that AI does not remove responsibility for accuracy. If disclosure or explanation receives a lower level of agreement, that difference could concern wording, conditions or the practical meaning of responsibility. It cannot be interpreted automatically as a hierarchy of ethical importance or evidence that respondents intend to conceal assistance.

The Group 4 instrument permits a focused examination. Its five authorship statements concern accuracy, verification, disclosure, explanation and understanding. Separate questions report verification and disclosure practices. A policy item asks whether respondents know if their institution has guidance. These statements address different objects and use different reference conditions. Analysing them separately allows the study to ask which principles are widely endorsed, whether similar aggregate proportions conceal differing individual responses, and how policy awareness is associated with a specific responsibility judgement.

---

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

[2] American Bar Association Standing Committee on Ethics and Professional Responsibility, ‘Generative Artificial Intelligence Tools’ (Formal Opinion 512, 29 July 2024).

A model that combines every authorship statement into one score could obscure these questions. It could also mix general duties with conditional disclosure and personal restraint. The present article therefore treats perceived responsibility as a set of specified judgements. The term is retained as the substantive topic, while the statistical analyses identify the particular statement being described or modelled.

Theory and research on academic integrity give considerable attention to contextual norms and institutional arrangements. McCabe and Treviño's study of honour codes is an important empirical contribution to that tradition.[3] More recent work examines generative-AI policies, student perceptions and assessment frameworks.[4] These contributions motivate a precise question for legal writing: whether endorsement of a duty and reporting a corresponding practice align within the same respondents, rather than merely appearing common in aggregate.

The contribution is deliberately narrower than a claim to identify the determinants of integrity across Bangladesh. The database includes 245 law students alongside graduates, advocates or legal professionals, and other respondents. The main analysis describes all 385 records, with the recorded law-student subset examined separately. It does not assume that all participants belong to university law programmes, have completed an AI-assisted assignment or have experienced a particular institutional policy.

This scope distinguishes an exploratory analysis of recorded judgements from a national estimate of student compliance. It also makes the proposed policy relationship testable without attributing intentions, enforcement experiences or observed conduct that the questionnaire does not record. No claim of national novelty is made on the basis of an assumed absence of prior studies.

The general objective is to examine perceived responsibility for AI-assisted legal writing and its relationship with reported practices in the supplied Bangladesh sample. Four questions organise the study. How frequently are the five responsibility statements endorsed? How do verification and disclosure duties align with corresponding practice reports within respondents? How are selected responsibility judgements associated on the original ordered scales? Is perceived disclosure responsibility associated with policy awareness, reported disclosure and accuracy responsibility after limited adjustment?

The analysis contributes full response distributions and within-person comparisons, alongside an explicit distinction between duty, practice and policy knowledge. It also reports the limits of the available measurement, rather than imposing a latent mediation or moderation model on

[3] Donald L McCabe and Linda Klebe Treviño, 'Academic Dishonesty: Honor Codes and Other Contextual Influences' (1993) 64 Journal of Higher Education 522.

[4] Hui Wang and others, 'Generative AI in Higher Education: Seeing ChatGPT through Universities' Policies, Resources, and Guidelines' (2024) 7 Computers and Education: Artificial Intelligence 100326; Cecilia Ka Yuk Chan and Wenjie Hu, 'Students' Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education' (2023) 20 International Journal of Educational Technology in Higher Education 43; Mike Perkins and others, 'The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment' (2024) 21(6) Journal of University Teaching and Learning Practice, doi:10.53761/q3azde36.

heterogeneous statements. These distinctions can inform the development of tasks that assess verification and explanation directly.

Section 2 develops the conceptual and theoretical foundation. Section 3 explains the records, measurements and analytical procedures. Results are presented in Section 4, followed by discussion in Section 5 and implications in Section 6. Section 7 sets out the contribution, limitations and future research. Section 8 records the authors' declarations, including the absence of formal ethical approval and the participant safeguards followed.

2. Literature Review and Theoretical Foundation

The theoretical foundation separates perceived duties, reports of action and the institutional conditions surrounding both. The Theory of Planned Behavior provides a language for distinguishing attitudes, intentions, control and conduct. Academic-integrity scholarship supplies the educational context. Neither framework warrants treating every favourable response as a measure of the same mechanism.

2.1 Responsibility for the final text

Responsibility for AI-assisted writing can include checking accuracy, assessing sources, explaining reasoning and acknowledging assistance where required. These activities are connected, but they are not interchangeable. A disclosure identifies something about the production process; it does not establish that the resulting argument is sound. A verified citation establishes that an authority exists or supports a proposition; it does not show that the writer can explain the reasoning. The present analysis preserves those distinctions in its interpretation of the five authorship items.

The legal-writing setting makes source verification especially relevant. Dahl and colleagues' study demonstrates why the apparent fluency of an answer cannot substitute for checking legal facts.[5] Comparative ABA guidance addresses human professional responsibilities in AI use, but the application of legal duties remains jurisdictional and task dependent.[6] In this article, perceived authorship responsibility is a questionnaire topic, not a determination of professional liability. A respondent's answer does not establish a breach, a sanction or the correctness of a legal argument.

The study consequently avoids defining responsibility as unlimited legal liability for every aspect of a submission. The measures ask about specific duties and personal judgements. That wording supports investigation of how respondents understand authorship, while leaving the applicable academic rule and legal consequence to separate analysis of the actual setting.

2.2 Disclosure and permitted assistance

Disclosure concerns identifying assistance in accordance with the rules governing a task. Its appropriate form may depend on whether a tool was used for language correction, idea generation,

[5] Dahl and others (n 1).

[6] American Bar Association (n 2).

legal analysis or substantial drafting. Perkins discusses the academic-integrity implications of large language models in assessment, highlighting the need to consider the educational use rather than equate all assistance.[7] A survey about disclosure should therefore specify the conditions under which acknowledgement is required.

Both the current duty and practice statements contain that condition. AR3 says substantial contribution should be disclosed when required by academic rules; GAI3 says the respondent discloses or acknowledges assistance when required. Their comparison is substantively useful because both concern disclosure, but it does not establish that respondents applied the same rule or recalled the same assignment. An answer can also reflect hypothetical judgement rather than recent conduct if a person has not faced a relevant task.

Assessment frameworks offer ways to make these conditions explicit. The AI Assessment Scale links permitted AI use to learning objectives, and its pilot implementation examines how such an approach can operate in a particular institution.[8] A later refinement continues the work of specifying educational assessment conditions.[9] These frameworks are relevant examples, not evidence that the respondents' institutions used those arrangements or that they caused the current response patterns.

2.3 Verification and understanding

Verification can involve checking whether a case or statute exists, whether it applies in the relevant jurisdiction and whether it supports the claim being made. A respondent may endorse the duty to verify without using a reliable method. Conversely, someone may report checking claims without expressing the same strength of agreement with a general duty statement. The present survey does not score either respondent's method. It reports their answers about verification.

Understanding also extends beyond finding a source. A writer should be able to explain why a proposition matters and how an authority relates to the facts. The AR4 explanation statement and AR5 restraint statement identify parts of that responsibility, but they do not supply an oral or written performance test. Endorsing them cannot demonstrate independent reasoning, just as non-endorsement cannot establish inability to reason.

AI literacy offers a broader educational reference. Long and Magerko identify competencies for understanding and critically engaging with AI.[10] In legal education, those competencies can inform tasks that distinguish a generated answer from a justified argument. The common literacy block in

[7] Mike Perkins, 'Academic Integrity Considerations of AI Large Language Models in the Post-Pandemic Era: ChatGPT and Beyond' (2023) 20(2) Journal of University Teaching and Learning Practice, art 7.

[8] Perkins and others (n 4); Leon Furze and others, 'The AI Assessment Scale (AIAS) in Action: A Pilot Implementation of GenAI-Supported Assessment' (2024) 40(4) Australasian Journal of Educational Technology 38.

[9] Mike Perkins, Jasper Roe and Leon Furze, 'Reimagining the Artificial Intelligence Assessment Scale: A Refined Framework for Educational Assessment' (2025) 22(7) Journal of University Teaching and Learning Practice, doi:10.53761/rrm4y757.

[10] 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) 1.

this survey covers several different ideas, including error awareness, verification, confidentiality and oversight. It is not assumed to be a validated single scale or a direct measure of legal research competence.

2.4 Theory of planned behaviour

Ajzen's Theory of Planned Behavior relates intentions to attitudes, subjective norms and perceived behavioural control, with behaviour also depending on opportunities and control.[11] It is useful here principally because it requires distinctions among beliefs, intended action and conduct. The questionnaire's responsibility statements express judgements about duties or restraint, while the GAI items report general practices. Those are related theoretical domains, but neither automatically operationalises every component of the theory.

The instrument does not contain separate, task-matched measures of subjective norms, perceived control and prospective intention for a particular legal-writing submission. A broad responsibility score therefore cannot serve as proof of the complete theoretical sequence. Likewise, a cross-sectional association between perceived duty and reported action would not establish which preceded the other. The article uses the theory to define the measurements needed for future work rather than claim confirmation of an unmeasured model.

A stronger design could specify an assignment, identify its AI-use rule, measure the student's intention before submission and observe later disclosure or verification. It could also examine whether students have the resources needed to act, such as access to primary legal sources or a clear explanation of permitted use. Such a design would make the connection between theoretical constructs and actual decisions more explicit.

2.5 Academic integrity and institutional context

Academic integrity encompasses individual decisions and the educational conditions in which they are made. McCabe and Treviño's empirical work relates dishonesty to contextual features including honour codes and peer environments.[12] Bretag's handbook presents a wider interdisciplinary treatment of integrity in education.[13] These sources support examining institutional context, but they do not supply validation for a new questionnaire block merely because its items concern rules or ethical conduct.

Policy awareness also needs a precise definition. IPR1 asks whether respondents know if their institution has rules or guidance. It does not verify the content, clarity, accessibility or enforcement of a policy. The other IPR statements concern the perceived value of rules or training, an intended restriction and an attributed enforcement benefit. Aggregating them could blur knowledge of an existing policy with approval of possible governance measures.

---

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

[12] McCabe and Treviño (n 3).

[13] Tracey Bretag (ed), *Handbook of Academic Integrity* (Springer 2016).

Research examining university policies offers useful context for that distinction. Wang and colleagues study institutional policies, resources and guidelines, while UNESCO provides broader educational guidance.[14] Neither source establishes that an awareness item measures policy quality. The current analysis therefore uses IPR1 as perceived awareness only. It does not describe its coefficient as an enforcement effect or a moderator of compliance.

2.6 Fair assessment and the limits of detection

Responsibility also applies to how institutions evaluate allegations of improper assistance. A detection score does not itself reveal what a writer understood, which tool was used or whether a rule was breached. Weber-Wulff and colleagues tested detection tools under several conditions, identifying limitations relevant to educational use.[15] Liang and colleagues also investigated the risk of misclassifying non-native English writing.[16] These studies motivate cautious assessment of evidence rather than an assumption that detection can directly establish misconduct.

For legal education, an explanation of reasoning and a record of source checking may provide different evidence from a detector output. Such approaches should still be evaluated for fairness, accessibility and relevance to the learning objective. The current survey contains no detector result, submitted assignment or disciplinary finding, so it cannot estimate a misconduct rate or validate a detection method.

Student perspectives matter in designing that evaluation. Chan and Hu examine perceived benefits and challenges of generative AI in higher education.[17] Those perceptions are distinct from demonstrated ability or compliance. The present study similarly treats respondents' judgements as informative about how responsibility is understood, while avoiding an inference that favourable answers verify the quality of their writing.

2.7 Analytical questions and framework

The analysis develops three connected lines of inquiry. First, it compares endorsement of the specific responsibility statements, without interpreting differences as a direct ranking of ethical importance. Second, it examines content-matched duty and practice pairs for verification and disclosure. Similar aggregate percentages may coexist with many respondents answering the two items differently. Joint patterns are therefore necessary alongside marginal summaries.

Third, the exploratory ordinal model examines perceived disclosure responsibility in relation to policy awareness, reported disclosure and accuracy responsibility. The model does not treat reported practice as a causal mediator or identify the effect of introducing a policy. It describes

[14] Wang and others (n 4); Fengchun Miao and Wayne Holmes, *Guidance for Generative AI in Education and Research* (UNESCO 2023).

[15] Debora Weber-Wulff and others, 'Testing of Detection Tools for AI-Generated Text' (2023) 19 International Journal for Educational Integrity 26.

[16] Weixin Liang and others, 'GPT Detectors Are Biased against Non-Native English Writers' (2023) 4 Patterns 100779.

[17] Chan and Hu (n 4).

conditional associations among separately worded responses, with limited adjustment for AI use, training and recorded student status.

These questions were specified for this analysis after examining the available instrument and data structure. They are not presented as preregistered confirmatory hypotheses. Figure 1 maps the measured objects and identifies the outcomes that remain unobserved. The framework focuses the article on authorship responsibility while preserving the distinction between agreeing with a principle and demonstrating responsible work.

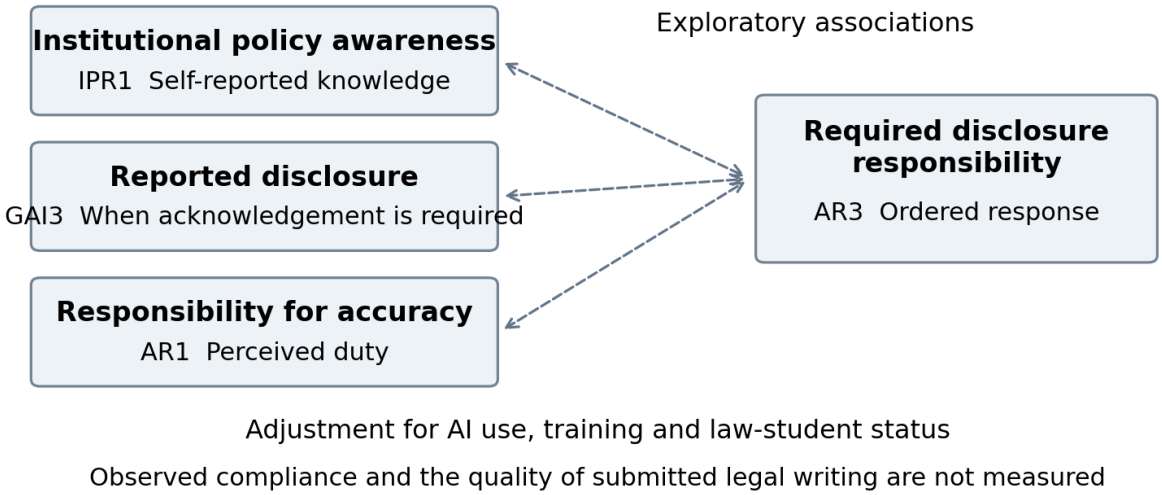


Figure 1. Exploratory framework for perceived disclosure responsibility

Dashed links denote associations. Duty–practice comparisons are analysed separately for verification and disclosure. No causal mediation, moderation or observed compliance pathway is represented.

3. Methodology

This cross-sectional study analyses the Group 4 numerical workbook supplied by the researchers as real survey data. Reporting distinguishes the dataset, the procedures implemented for this analysis and the ethical arrangements confirmed by the authors. It follows the transparency objective of observational reporting guidance, without treating a proposed sampling protocol as evidence of completed field procedures.[18]

3.1 Analytical sample and scope

There are 385 records, all coded as consenting adults with relevant AI awareness. Current status identifies 245 law students, 71 law graduates, 37 advocates or legal professionals and 32 other relevant respondents. The full sample supports a description of perceived student responsibility

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

across these recorded groups. The law-student subset is analysed separately; the other records are not relabelled as university students.

The unit of analysis is an individual recorded response. The workbook does not link responses to institutions, assignments, submitted texts or observed writing sessions. It therefore cannot verify current programme enrolment, exposure to a named AI policy or actual use of AI for a specific legal-writing task. The institutional referent is not independently verified, particularly for non-students. These boundaries determine the scope of both the title and the interpretation.

3.2 Sampling and provenance

Cochran's conventional large-population calculation using $Z = 1.96$, $p = 0.50$ and $e = 0.05$ yields 384.16, rounded to 385.[19] That calculation can motivate a planning target under simple-random-sampling assumptions. It does not establish that 428 people were approached, that every retained record arose from probability sampling or that a 5% national margin of error applies.

Fieldwork dates, invitation counts, institutions, recruitment channels and a response-rate denominator are not documented in the supplied materials. A separate pilot dataset, pilot feedback and revision history are also unavailable. No completed pilot, hybrid distribution protocol or specific recruitment city is inferred from general methodological descriptions. All 385 available records are retained; there is no reconstructed response rate or survey weight.

The Codebook labels the respondent sequence and contextual comments as synthetic. Those labels require explanation against original records but do not alone establish that numerical answers are simulated. Contextual text is excluded throughout. Identifiers distinguish rows and are not used as substantive variables. The analysis is conditional on the supplied numerical data and does not independently authenticate recruitment.

3.3 Item definitions and measurement decisions

The focal AR items use five ordered agreement categories. Table 1 identifies their content and the specific interpretation supported. The exact wording, together with the supporting disclosure, verification and policy items, is reproduced in Appendix A. The source is the supplied Codebook; no published scale-adaptation record is asserted without documentation.

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

Table 1. Authorship statements and their analytical interpretation

Item	Statement focus	Interpretation
AR1	Responsibility for accuracy	Perceived general duty; not an accuracy test
AR2	Duty to verify authorities and citations	Perceived verification duty; not observed checking
AR3	Disclosure when academic rules require it	Conditional responsibility judgement; ordinal model outcome
AR4	Explaining and defending submitted text	Perceived explanation duty; not assessed reasoning
AR5	Avoiding reliance on ununderstood writing	Conditional restraint; not an observed submission decision

Source: Group 4 workbook Codebook. Exact focal and supporting statements appear in Appendix A.

All four five-item blocks were examined diagnostically. Recomputed coefficient alpha is 0.205 for AIL, 0.301 for GAI, 0.225 for AR and 0.135 for IPR. These estimates do not support retaining the supplied composite-scale reliability claims. More fundamentally, the blocks contain different response objects. Alpha alone cannot establish dimensionality or content validity, and raising it by selective deletion would not establish an adoption or integrity construct.[20] The article therefore reports items separately and does not estimate latent mediation, moderation, composite reliability, AVE or HTMT.

3.4 Data screening

The focal AR items, supporting items and model covariates contain no missing values. Fourteen records lack the optional AI-tool response, which is not used in the focal analyses. Identifiers are unique. No exact duplicate is found after excluding identifier and contextual-text field. Eleven respondents give one constant answer across the AR block; none gives a constant response across all 20 Likert items.

Constant focal responses are retained because agreement across related duties can be substantive. Completion times and an objectively scored attention check are unavailable. Numerical screening therefore establishes recorded consistency rather than independent identity or attention. No imputation, weighting, outcome-driven deletion or undocumented recoding is performed.

3.5 Descriptive and paired analyses

Full response counts, means, standard deviations and endorsement proportions are reported. Endorsement combines agree and strongly agree; neutral remains visible in the full distributions. Pointwise Wilson intervals accompany endorsement proportions under conventional

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

assumptions.[21] These intervals do not resolve selection bias or establish national representativeness.

Four paired contrasts compare AR1 with AR3, AR2 with AR4, AR2 with reported verification GAI4, and AR3 with reported disclosure GAI3. Exact McNemar tests use discordant pairs, with Holm correction across the four tests.[22] Pointwise percentile intervals use 10,000 whole-respondent bootstrap resamples, preserving the pairing.[23] Joint counts for the matched duty–practice pairs report both endorsed, duty only, practice only and neither. Different wording and reference conditions mean that discordance is not classified as misconduct.

Seven selected Spearman correlations describe the ordered relationships: AR1 with each other AR item, the two matched duty–practice pairs, and IPR1 with AR3. Pointwise intervals use 3,000 respondent bootstrap resamples; Holm correction covers the seven tests as one family. Effect sizes and uncertainty are interpreted together.[24]

3.6 Exploratory ordinal model and sensitivity

The proportional-odds model uses perceived required-disclosure responsibility, AR3, as its outcome. Its four observed categories are disagree, neutral, agree and strongly agree. Predictors are IPR1 policy awareness, GAI3 reported disclosure and AR1 accuracy responsibility, each scored per one-point increase, plus weekly-or-more AI use, any formal training and law-student status. The reference categories for the binary covariates are their complements. This is a conditional association model, not a causal adjustment design.[25]

Maximum-likelihood estimates use observed-information Wald intervals. Holm correction covers the three focal coefficients separately from the correlation and paired-test families. A cluster-robust Wald comparison of threshold-specific logistic slopes examines the common-slope assumption while accounting for repeated thresholds within respondents.[26] Robust binary models

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

[22] 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.

[23] Bradley Efron and Robert J Tibshirani, *An Introduction to the Bootstrap* (Chapman & Hall 1993).

[24] Ronald L Wasserstein and Nicole A Lazar, 'The ASA Statement on p-Values: Context, Process, and Purpose' (2016) 70 American Statistician 129.

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

[26] Rollin Brant, 'Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression' (1990) 46 Biometrics 1171; Kung-Yee Liang and Scott L Zeger, 'Longitudinal Data Analysis Using Generalized Linear Models' (1986) 73 Biometrika 13.

at the neutral-or-higher, agree-or-higher and strongly-agree thresholds provide additional sensitivity summaries.[27]

A student-restricted ordinal model omits status, which is constant within the subset. Variance inflation factors assess collinearity, and two optimisation methods check numerical stability. These analyses are exploratory, and separate multiplicity families do not provide a single article-wide error guarantee. Computation used Python with pandas, NumPy, SciPy and statsmodels.

3.7 Ethics and participant protection

Ethical approval for this study was obtained from the Center for Academic & Professional Career Development and Research (CAPCDR), an external institutional body, under approval reference CAPCDR: SMUCT/LLB/SS-0020261012-03. 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.

The authors confirm that informed consent preceded questionnaire completion, participation was voluntary and anonymous, and respondents could discontinue. All analysed records contain affirmative consent and adult-eligibility codes. Reporting is aggregate, contextual comments are excluded and no attempt is made to identify individuals. Discontinuation is not described as a promise to trace and remove an already anonymous individual submission.

4. Results

Results distinguish the frequency of endorsing a principle from agreement between related answers. They concern recorded perceptions and reports, with no observation of a submitted assignment, source-checking procedure or academic-integrity decision. Percentages use the full sample unless another denominator is stated.

4.1 Respondent profile and analytical coverage

Table 2 describes recorded status and training. These categories establish the mixed composition of the full sample and the separately identifiable student subset. They do not identify particular institutions or legal-writing courses.

[27] David W Hosmer Jr, Stanley Lemeshow and Rodney X Sturdivant, *Applied Logistic Regression* (3rd edn, John Wiley & Sons 2013); Halbert White, 'A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity' (1980) 48 Econometrica 817.

Table 2. Recorded respondent status and training

Measure	Category	Count (%)
Current status	Law student	245 (63.6%)
	Law graduate	71 (18.4%)
	Advocate or legal professional	37 (9.6%)
	Other relevant respondent	32 (8.3%)
AI training	No	202 (52.5%)
	Yes, short orientation	127 (33.0%)
	Yes, course or certified training	56 (14.5%)

N = 385. The recorded law-student subset contains 245 observations. Other records are not classified as current university law students; named courses and institutions are not identified.

Law students account for 245 records (63.6%), law graduates for 71 (18.4%), advocates or legal professionals for 37 (9.6%), and other respondents for 32 (8.3%). The largest age category is 21–23 years, with 154 respondents (40.0%), followed by 24–26 years with 89 (23.1%). Gender is recorded as male for 202 respondents (52.5%), female for 171 (44.4%), prefer not to say for nine and other for three.

Residence or work location is Dhaka for 147 records (38.2%), a district town for 108 (28.1%), another city corporation for 72 (18.7%), an upazila or rural area for 50 (13.0%) and other for eight. These labels do not establish that recruitment occurred through those locations. Weekly-or-more AI use is reported by 272 respondents (70.6%). Fourteen report never using generative AI, which is compatible with eligibility based on having heard of AI. Any formal training is recorded for 183 respondents (47.5%). All main-model variables are complete.

4.2 Distributions of perceived responsibility

Table 3 reports all five AR distributions. Neutral responses remain distinct so that lower endorsement is not equated with opposition to a duty. Means and standard deviations are descriptive summaries of the ordered scale.

Table 3. Distributions of perceived authorship responsibility

Item	Counts 1 / 2 / 3 / 4 / 5	Mean	SD	Endorsed n (%)	95% CI (%)
AR1	0 / 5 / 35 / 199 / 146	4.26	0.67	345 (89.6%)	86.2 to 92.3
AR2	0 / 3 / 54 / 176 / 152	4.24	0.71	328 (85.2%)	81.3 to 88.4
AR3	0 / 18 / 110 / 182 / 75	3.82	0.80	257 (66.8%)	61.9 to 71.3
AR4	1 / 11 / 112 / 187 / 74	3.84	0.77	261 (67.8%)	63.0 to 72.3
AR5	2 / 25 / 126 / 176 / 56	3.67	0.82	232 (60.3%)	55.3 to 65.0

N = 385. Categories 1–5 run from strongly disagree to strongly agree. Endorsement combines agree and strongly agree. Wilson intervals are pointwise. Means and SDs are descriptive summaries.

Responsibility for accuracy, AR1, is endorsed by 345 respondents (89.6%), with a mean of 4.26 and standard deviation of 0.67. The duty to verify authorities and citations, AR2, is endorsed by 328 (85.2%; mean 4.24). Required disclosure, AR3, is endorsed by 257 (66.8%; mean 3.82). Explaining and defending submitted text, AR4, is endorsed by 261 (67.8%; mean 3.84), while avoiding reliance on writing that is not understood, AR5, is endorsed by 232 (60.3%; mean 3.67).

The neutral counts differ substantially: 35 for accuracy, 54 for verification duty, 110 for disclosure, 112 for explanation and 126 for the restraint statement. Strong agreement occurs in 146 accuracy responses and 152 verification-duty responses, compared with 75 for disclosure, 74 for explanation and 56 for restraint. These distributions show differentiated responses rather than a uniform position across every responsibility item.

Supporting-item endorsement is 247 (64.2%) for reported disclosure, 325 (84.4%) for reported verification and 233 (60.5%) for policy awareness. These reports are not treated as audited compliance or verified knowledge of an institutional document.

4.3 Paired differences between statements

Table 4 reports four selected contrasts, retaining within-respondent pairing. The first two compare different responsibility statements; the latter two compare a duty with a content-matched practice report. The resulting differences have different substantive interpretations.

Table 4. Paired endorsement differences

Contrast	First only / second only	Difference pp	95% CI pp	Holm p
AR1 - AR3	115 / 27	22.9	17.4 to 28.6	<0.001
AR2 - AR4	105 / 38	17.4	11.4 to 23.4	<0.001
AR2 - GAI4	49 / 46	0.8	-4.2 to 5.7	0.980
AR3 - GAI3	90 / 80	2.6	-4.2 to 9.1	0.980

N = 385; pp = percentage points. Exact McNemar tests are Holm-adjusted across four comparisons. Pointwise percentile intervals use 10,000 respondent bootstrap resamples. Different item content and reference conditions limit interpretation.

Accuracy responsibility exceeds required-disclosure endorsement by 22.9 percentage points (95% bootstrap interval 17.4–28.6; Holm p below 0.001). Verification duty exceeds the explanation statement by 17.4 points (11.4–23.4; Holm p below 0.001). These contrasts indicate different response proportions; they do not establish that respondents consider one duty more important under otherwise identical conditions.

Verification duty exceeds reported verification by only 0.8 percentage points, with an interval of -4.2 to 5.7 points. Required-disclosure endorsement exceeds reported disclosure by 2.6 points, with an interval of -4.2 to 9.1 points. Both matched comparisons have Holm p = 0.980. The results provide no clear difference in their marginal endorsement proportions, but do not establish equivalence or close alignment within individuals.

4.4 Joint patterns in duty and practice reports

Table 5 examines that within-person alignment. Each panel partitions the full sample into four exhaustive response combinations. A non-endorsed item can be neutral or disagreeing, so single-item categories are not described as rejection of an ethical principle.

Table 5. Joint responses for matched duty and practice statements

Item pair	Response combination	Count (%)
AR2 / GAI4	Both endorsed	279 (72.5%)
	Duty only	49 (12.7%)
	Practice only	46 (11.9%)
	Neither	11 (2.9%)
AR3 / GAI3	Both endorsed	167 (43.4%)
	Duty only	90 (23.4%)
	Practice only	80 (20.8%)
	Neither	48 (12.5%)

N = 385. AR2/GAI4 concerns verification; AR3/GAI3 concerns disclosure. Non-endorsement includes neutral responses. Discordance is a response pattern, not an observed breach of an academic rule.

Verification duty and reported verification are both endorsed in 279 records (72.5%). Forty-nine respondents endorse the duty alone, 46 report the practice alone and 11 endorse neither. Thus 95 records (24.7%) differ in binary endorsement despite the closely matched marginal percentages. The almost balanced single-item counts explain why the paired difference is small.

Required disclosure and reported disclosure are both endorsed in 167 records (43.4%). Ninety respondents endorse the duty alone, 80 report the practice alone and 48 endorse neither. The two items differ in binary endorsement in 170 records (44.2%). That pattern describes answers under their respective wordings. It does not identify concealment, a breach of a known rule or a verified intention–behaviour discrepancy.

4.5 Ordinal associations

Table 6 uses all ordered categories to describe seven selected relationships. The correction family includes every correlation listed, and the bootstrap intervals are pointwise rather than multiplicity-adjusted.

Table 6. Selected ordinal associations

Item pair	Spearman ρ	95% CI	Raw p	Holm p
AR1 / AR2	0.037	-0.068 to 0.141	0.467	1.000
AR1 / AR3	0.102	-0.002 to 0.201	0.046	0.322
AR1 / AR4	0.076	-0.022 to 0.171	0.137	0.684
AR1 / AR5	0.043	-0.057 to 0.141	0.402	1.000
AR2 / GAI4	0.012	-0.088 to 0.112	0.816	1.000
AR3 / GAI3	0.075	-0.027 to 0.176	0.144	0.684
IPR1 / AR3	0.092	-0.010 to 0.189	0.072	0.430

N = 385. Holm correction covers seven correlations. Pointwise bootstrap intervals use 3,000 whole-respondent resamples. These estimates describe ordered questionnaire responses.

Accuracy responsibility has small correlations with the other AR statements: 0.037 with verification duty, 0.102 with disclosure, 0.076 with explanation and 0.043 with restraint. The two matched duty–practice correlations are 0.012 for verification and 0.075 for disclosure. Policy awareness correlates with required-disclosure responsibility at 0.092.

All seven Holm-adjusted p-values exceed 0.05, and the reported bootstrap intervals include zero. The largest estimate, accuracy with disclosure, has raw p = 0.046 but Holm p = 0.322. Its bootstrap interval is approximately -0.002 to 0.201. The selected associations are therefore small and uncertain. These findings do not prove that the principles are unrelated in the population or that institutional guidance has no effect.

4.6 Adjusted model and sensitivity analyses

Table 7 reports the ordinal model of required-disclosure responsibility. A positive odds ratio indicates higher odds of being above a response threshold, conditional on the included variables. It is not a change in disclosure probability or an intervention effect.

Table 7. Ordinal model of required disclosure responsibility

Predictor	aOR	95% CI	Raw p	Holm p
Policy awareness IPR1	1.20	0.94 to 1.53	0.137	0.410
Reported disclosure GAI3	1.16	0.92 to 1.48	0.210	0.410
Accuracy responsibility AR1	1.22	0.92 to 1.61	0.167	0.410
Weekly-or-more AI use	1.56	1.02 to 2.38	0.040	—
Any formal AI training	1.19	0.81 to 1.73	0.378	—
Law-student status	0.98	0.66 to 1.45	0.914	—

N = 385. AR3 categories: disagree 18, neutral 110, agree 182 and strongly agree 75. Focal estimates are per one-point increase. Reference categories: less frequent AI use, no training and non-student status. Observed-information intervals are pointwise. Holm correction covers the three focal coefficients; covariate estimates are descriptive. aOR means adjusted odds ratio.

Policy awareness has an adjusted odds ratio of 1.20 per one-point increase (95% interval 0.94–1.53; raw p = 0.137), reported disclosure 1.16 (0.92–1.48; raw p = 0.210), and accuracy responsibility 1.22 (0.92–1.61; raw p = 0.167). All three focal Holm p-values are 0.410. The estimates are compatible with modest positive associations but also with weaker or absent associations under the model.

Weekly-or-more AI use has a descriptive estimate of 1.56 (1.02–2.38), any training 1.19 (0.81–1.73), and law-student status 0.98 (0.66–1.45). These covariate comparisons are not multiplicity-adjusted confirmatory findings. Variance inflation factors range from 1.008 to 1.034. The model converges, and the two optimisers agree to within 0.001. The global threshold diagnostic gives $\chi^2 = 7.21$ on 12 degrees of freedom, $p = 0.844$, providing no evidence against common slopes without proving that restriction.

Table 8 reports the student subset. Its denominator is 245, while the database and principal analysis contain 385 records. These are overlapping analyses rather than independent replications.

Table 8. Responsibility and supporting items in the student subset

Item	Full sample n (%)	Law students n (%)
AR1	345 (89.6%)	221 (90.2%)
AR2	328 (85.2%)	202 (82.4%)
AR3	257 (66.8%)	164 (66.9%)
AR4	261 (67.8%)	166 (67.8%)
AR5	232 (60.3%)	153 (62.4%)
GAI3	247 (64.2%)	160 (65.3%)
GAI4	325 (84.4%)	204 (83.3%)
IPR1	233 (60.5%)	151 (61.6%)

Denominators: full sample 385; law students 245. The subset overlaps the full sample. The student-only ordinal model removes constant student status; its focal estimates are reported below. No independent replication or between-population difference is implied.

In the student-only ordinal model, the policy-awareness odds ratio is 1.13 (0.84–1.52), reported disclosure 1.16 (0.86–1.56), and accuracy responsibility 1.24 (0.87–1.76). These intervals are pointwise and span one. The restriction therefore does not resolve the uncertainty in the focal conditional associations or establish a difference between students and other respondents.

5. Discussion

The principal finding is a distinction between common endorsement of responsibility and limited alignment among the specific statements. Accuracy and verification receive high agreement, while disclosure, explanation and restraint attract more neutral responses. Similar marginal proportions for a duty and a reported practice also conceal substantial differences within respondents. The discussion interprets those patterns without converting them into allegations of misconduct or causal policy effects.

5.1 Broad responsibility and its specific expressions

Accuracy responsibility is widely endorsed, but it has only small associations with the other AR statements. A broad principle can be easy to endorse while its applications depend on context. Required disclosure refers to rules, explanation refers to a submitted text, and restraint refers to the writer's understanding. Those conditions can change an answer even where the underlying commitment to responsible writing remains important.

The findings therefore support a differentiated educational question: what must a writer do to discharge a particular responsibility? Legal-hallucination research provides a reason to assess source checking, while comparative professional guidance addresses duties in specific practice

contexts.[28] Neither source makes a general responsibility statement a sufficient measure of performance. The present data likewise cannot demonstrate that people who endorse accuracy can actually verify an authority or explain an argument.

This distinction is useful for assessment design. An educator can ask a learner to locate the authority supporting a proposition, identify its jurisdiction and explain its relevance. Such a task would produce evidence different from agreement with a duty. A future study could then examine whether perceived responsibility predicts performance, instead of assuming that the two are equivalent.

5.2 Similar percentages do not establish individual alignment

The matched verification items have nearly equal endorsement proportions, but 95 respondents differ in binary endorsement. The disclosure items have a similarly small marginal difference alongside 170 differing records. A comparison of the overall percentages alone would miss these combinations because duty-only and practice-only answers largely offset one another.

The combinations should not be interpreted as direct evidence that people fail to implement an intention. The duty and practice statements have different subjects and reference conditions, and the survey does not record whether disclosure was required in a particular assignment. An individual may also interpret acknowledging a tool differently from disclosing a substantial contribution. Interviews and task-specific follow-up could examine these possibilities; the current records cannot determine which applies.

The Theory of Planned Behavior helps clarify why this matters. A theoretical account of intention and conduct requires action-matched measures, including opportunities and control.[29] The present items are not a complete implementation of that design. Their contribution is to show what must be specified before a duty–practice relationship can be evaluated as a behavioural mechanism.

5.3 Disclosure responsibility and policy awareness

The ordinal policy-awareness estimate is positive but uncertain after adjustment. The item asks whether respondents know if their institution has rules; it does not independently establish that the rules are clear, accessible or consistently applied. The result therefore cannot show that a policy intervention succeeded or failed. It also cannot identify moderation by enforcement, which was not operationalised as an interaction between measured constructs.

Research on university policies distinguishes institutional documents and resources from individual responses to them.[30] The AI Assessment Scale and its subsequent development similarly provide examples of linking permitted assistance to assessment objectives.[31] They

[28] Dahl and others (n 1); American Bar Association (n 2).

[29] Ajzen (n 11).

[30] Wang and others (n 4).

[31] Perkins and others (n 4); Perkins, Roe and Furze (n 9).

motivate clearer task conditions, but the current data do not test whether adopting those frameworks changes disclosure responsibility.

A stronger evaluation would document the actual policy, assess comprehension and observe a relevant submission. Students could be given examples of permitted assistance and asked to prepare an appropriate acknowledgement. The evaluation could then distinguish awareness that a rule exists from understanding the rule and applying it correctly. The current single-item association cannot supply those distinctions, which is why an uncertain coefficient should not be turned into a broad judgement about policy value.

5.4 Integrity as an institutional and educational practice

Academic-integrity research places individual decisions within educational context. McCabe and Treviño's work provides evidence on contextual correlates, while Bretag's handbook addresses the wider field.[32] Those contributions support attention to guidance, expectations and institutional processes. They do not justify assuming that every favourable answer measures one shared integrity disposition.

The current block reliabilities reinforce this measurement concern. Low alpha does not prove that a responsibility topic is unimportant, nor does it establish why the items differ. The substantive issue is that an index would combine duty statements, action reports and perceived benefits of governance. An internally consistent but conceptually inappropriate score would also be problematic. Sijtsma's analysis of alpha supports treating it as limited evidence rather than a general certificate of measurement quality.[33]

The appropriate next step is a better specified instrument. Researchers can define each responsibility, use multiple content-relevant indicators where appropriate and assess whether those indicators support the proposed score. They should document item development and evaluate knowledge and action separately. The present results support that work; they do not warrant retaining a structural model merely to satisfy a preferred reporting template.

5.5 Fairness in evaluating AI assistance

A responsibility-based approach should apply to institutional assessment as well as student writing. Detection tools can provide a signal for further inquiry, but a detector output cannot itself establish the circumstances in which a text was produced. Empirical work on detection performance and possible language-related misclassification supplies a reason to examine the evidence and the decision process carefully.[34]

This has particular relevance where students use English as an additional language or use permitted language support. The current study does not measure language background or detector outcomes,

[32] McCabe and Treviño (n 3); Bretag (n 13).

[33] Sijtsma (n 20).

[34] Weber-Wulff and others (n 15); Liang and others (n 16).

so it does not estimate a local false-positive rate. It instead supports a proposed evaluation process that examines sources, explanations and the stated assignment rules. Any such process should allow the writer to explain the work and should distinguish assistance permitted by the task from a substantiated breach.

Student perceptions also need to be heard without treating them as performance evidence. Chan and Hu's research identifies perceived opportunities and challenges, while UNESCO's guidance addresses human capacity and educational governance.[35] In the present setting, those perspectives inform recommendations for transparent assessment. They do not supply a causal explanation for the observed responses.

5.6 Integrated interpretation and sample scope

The student subset generally retains the ordering seen in the full sample, with high endorsement of accuracy and verification and lower endorsement of the other responsibility statements. Its focal model estimates remain uncertain. This provides a useful scope check but does not make the mixed sample representative of all Bangladesh law students or establish that student and professional populations are identical.

The overall interpretation is that responsibility has several specified expressions whose prevalence and association should be reported separately. General endorsement is not proof of demonstrated compliance, and near-equal aggregate duty and practice percentages are not proof of close alignment within respondents. Conversely, differing answers are not evidence of dishonesty without a matched task and a known rule. These distinctions provide a more defensible basis for legal-writing education than a claim that all hypothesised structural relationships have been validated.

6. Theoretical and Practical Implications

The implications concern measurement, instruction and institutional evaluation. They identify what should be tested next. The study did not deliver or evaluate a course, introduce a policy or observe a legal-writing submission, so it cannot promise that a particular intervention will improve compliance.

6.1 Implications for theory and measurement

The first implication is to distinguish a general responsibility principle from its task-specific expression. Accuracy, disclosure and explanation may be related, but their empirical relationship cannot be assumed from the word responsibility. A future model should state whether it concerns a perceived duty, prospective intention, reported practice or observed performance. Each requires measures matched to its referent and time frame.

[35] Chan and Hu (n 4); Miao and Holmes (n 14).

Second, aggregate and within-person evidence answer different questions. A study can find that a duty and a practice are equally common while many respondents endorse only one. Reporting both marginal and joint distributions avoids inferring individual consistency from a population summary. Direct observation would then be needed to determine how these answers relate to actual conduct.

Third, theoretical integration should follow adequate operationalisation. TPB cannot be confirmed without the relevant intention, norm and control constructs. Institutional-policy awareness cannot stand in for policy clarity or enforcement. Researchers can use academic-integrity scholarship to guide those distinctions, while developing a measurement model that reflects the specific educational decision. The present study contributes the distinctions and descriptive evidence needed for that design, rather than a validated latent causal pathway.

6.2 Implications for legal education

Law faculties can translate general responsibility into observable tasks. A source-verification exercise could require a student to locate each cited authority, check the proposition it supports and identify any jurisdictional limitation. A short explanation of the reasoning could assess understanding separately from the accuracy of a generated draft. Fictional or public materials can be used so that assessment does not expose confidential client information.

Disclosure instruction should be equally concrete. An assignment can identify the permitted forms of assistance and require an acknowledgement proportionate to the contribution. Examples can distinguish language editing, brainstorming and substantive drafting without assuming that the same rule applies to every task. Framework and pilot work on the AI Assessment Scale provide possible starting points for local evaluation, rather than evidence that one format will work in every course.[36]

Evaluation should record whether students understand the rule, whether their acknowledgement is appropriate and whether their final reasoning is defensible. A change in agreement with a general statement should not be the only outcome. Feedback can address errors in source use or disclosure without treating uncertainty as proof of bad intent. This follows the study's distinction between responsibility judgements and demonstrated work.

6.3 Implications for institutional governance

Institutions can examine whether AI guidance is findable, understandable and aligned with assessment objectives. A useful policy-awareness measure would include locating a rule and applying it to a scenario, rather than only reporting that guidance exists. Training evaluations should document content and assess the relevant task before and after instruction, with a comparison design where feasible.

[36] Perkins and others (n 4); Furze and others (n 8).

Governance should also provide a fair process for reviewing suspected misuse. Evidence from detection studies supports caution about relying on automated classifications alone.[37] Source records, the writer's explanation and the applicable rule can inform a more specific inquiry. These recommendations concern educational procedure and remain subject to local evaluation; the survey does not establish a disciplinary standard or a Bangladesh-wide legal duty. Evaluation should also consider access to primary sources, accessible formats and the time required to complete a verification task. Students should be able to identify when further guidance is needed. A fair assessment should distinguish that justified request for support from an unsupported assumption that the student has not taken responsibility.

7. Conclusion Limitations and Future Research

The study describes perceived responsibility and reported practices in AI-assisted legal writing across 385 supplied records, including a separate analysis of 245 law students. Accuracy responsibility and verification duty receive high endorsement. Disclosure, explanation and restraint receive lower endorsement and more neutral responses. Content-matched duty and practice statements have similar marginal percentages but substantial within-person differences. Selected ordinal associations and the focal adjusted disclosure-responsibility estimates are small or uncertain.

The contribution is a differentiated empirical account of responsibility, not confirmation of a policy-to-integrity mechanism. It shows why favourable self-reports cannot substitute for observing source verification, disclosure or understanding. It also shows why aggregate similarity and individual alignment should be examined separately. These distinctions allow legal education to ask more precise questions about the work a student can explain and defend.

The design has important limits. It is cross-sectional and based on study-specific single items. There is no submitted text, objective legal task, policy-document audit or observed disciplinary outcome. The sample includes students and other respondents, with recruitment, dates and response-rate information undocumented. The Codebook's synthetic identifier and contextual-text labels require explanation against original records; the numerical data were supplied as real responses, and comments were excluded. Statistical adjustment cannot resolve these provenance and selection questions.

The ordinal model also assumes a constant log-odds association for each one-point predictor increase and common slopes across outcome thresholds. The threshold diagnostic did not identify a departure, but it does not validate every specification. The student restriction is an overlapping sensitivity analysis, and all inference remains exploratory. Uncertain estimates should therefore motivate better evidence rather than be interpreted as proof that guidance or responsibility has no practical importance.

[37] Weber-Wulff and others (n 15); Liang and others (n 16).

Future studies should document eligibility, recruitment and field procedures, then specify a legal-writing task and its permitted AI assistance. Cognitive interviews can examine how students interpret verification, substantial contribution and explanation. Multiple indicators can be developed for any proposed construct, with validation conducted before substantive paths are fitted. Objective tasks should remain distinct from perceptions so that calibration can be assessed.

A longitudinal or experimental design could then examine whether instruction changes an appropriate disclosure, the quality of source checking or the ability to defend an argument. It should record the rule, the assistance used and the support available, while protecting participants' privacy. Such evidence would allow a stronger assessment of how educational guidance connects to responsible work. The present findings provide a transparent starting point and specify the boundaries that a subsequent study must address.

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-0020261012-03. 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

The authors confirm informed consent before completion, voluntary participation, anonymity and the opportunity to discontinue. All analysed records contain affirmative consent and adult-eligibility responses.

Data availability

De-identified numerical data and analytical specifications may be requested from the corresponding author, subject to confidentiality and applicable institutional requirements. Contextual comments were excluded.

Funding and competing interests

The authors report no external funding and no competing interests.

Author contributions

Mst. Sadia Rahman Arpita: Conceptualisation; study materials; data preparation and curation; manuscript development. Prof. Dr Kazi Abdul Mannan: Methodological guidance; supervision; interpretation of findings; critical review and editing.

Generative AI disclosure

Generative AI assisted literature searching, drafting, language editing, organisation and analytical-code development. Estimates were calculated programmatically from the supplied numerical records.

Table B1. Disclosure responsibility across response thresholds

Threshold	Predictor	Events n	aOR	95% CI
Neutral or higher	Policy awareness IPR1	367	1.70	0.88 to 3.27
	Reported disclosure GAI3		1.27	0.66 to 2.44
	Accuracy responsibility AR1		1.01	0.50 to 2.06
Agree or higher	Policy awareness IPR1	257	1.15	0.88 to 1.50
	Reported disclosure GAI3		1.15	0.88 to 1.50
	Accuracy responsibility AR1		1.15	0.85 to 1.56
Strongly agree	Policy awareness IPR1	75	1.19	0.85 to 1.67
	Reported disclosure GAI3		1.19	0.84 to 1.68
	Accuracy responsibility AR1		1.44	0.94 to 2.21

N = 385 for each threshold. Models use the same six predictors as the ordinal model and robust sandwich intervals. Intervals are pointwise and exploratory; no separate threshold-level significance claims are made. Neutral-or-higher includes neutral responses and has only 18 observations below its threshold.

The focal intervals span one at each threshold. The sparse lower response group contributes to wide intervals at the neutral-or-higher cut-off; these summaries do not prove equality of threshold-specific associations.

Bibliography

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

American Bar Association Standing Committee on Ethics and Professional Responsibility, 'Generative Artificial Intelligence Tools' (Formal Opinion 512, 29 July 2024)

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

Bretag T (ed), *Handbook of Academic Integrity* (Springer 2016)

Chan CKY and Hu W, 'Students' Voices on Generative AI: Perceptions, Benefits, and Challenges in Higher Education' (2023) 20 International Journal of Educational Technology in Higher Education 43

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

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)

Furze L and others, ‘The AI Assessment Scale (AIAS) in Action: A Pilot Implementation of GenAI-Supported Assessment’ (2024) 40(4) Australasian Journal of Educational Technology 38

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

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

Liang KY and Zeger SL, ‘Longitudinal Data Analysis Using Generalized Linear Models’ (1986) 73 Biometrika 13

Liang W and others, ‘GPT Detectors Are Biased against Non-Native English Writers’ (2023) 4 Patterns 100779

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) 1

McCabe DL and Treviño LK, ‘Academic Dishonesty: Honor Codes and Other Contextual Influences’ (1993) 64 Journal of Higher Education 522

McCullagh P, ‘Regression Models for Ordinal Data’ (1980) 42 Journal of the Royal Statistical Society Series B 109

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

Miao F and Holmes W, *Guidance for Generative AI in Education and Research* (UNESCO 2023)

Perkins M, ‘Academic Integrity Considerations of AI Large Language Models in the Post-Pandemic Era: ChatGPT and Beyond’ (2023) 20(2) Journal of University Teaching and Learning Practice, art 7

Perkins M and others, ‘The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment’ (2024) 21(6) Journal of University Teaching and Learning Practice, doi:10.53761/q3azde36

Perkins M, Roe J and Furze L, ‘Reimagining the Artificial Intelligence Assessment Scale: A Refined Framework for Educational Assessment’ (2025) 22(7) Journal of University Teaching and Learning Practice, doi:10.53761/rrm4y757

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

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

Wang H and others, ‘Generative AI in Higher Education: Seeing ChatGPT through Universities’ Policies, Resources, and Guidelines’ (2024) 7 Computers and Education: Artificial Intelligence 100326

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

Weber-Wulff D and others, ‘Testing of Detection Tools for AI-Generated Text’ (2023) 19 International Journal for Educational Integrity 26
White H, ‘A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity’ (1980) 48 Econometrica 817
Wilson EB, ‘Probable Inference, the Law of Succession, and Statistical Inference’ (1927) 22 Journal of the American Statistical Association 209

Appendix A Questionnaire wording

Table A1 reproduces the statements used in the focal analyses, including the supporting verification, disclosure and policy-awareness items.

Table A1. Exact wording of analysed statements

Code	Questionnaire statement
AR1	The student remains responsible for the accuracy of AI-assisted legal writing.
AR2	Using AI does not remove the duty to verify legal authorities and citations.
AR3	Substantial AI contribution should be disclosed when required by academic rules.
AR4	I should be able to explain and defend any AI-assisted text I submit.
AR5	I would not rely on AI-generated legal writing that I do not understand.
GAI3	I disclose or acknowledge AI assistance when required.
GAI4	I verify AI-generated citations and legal claims before using them in academic work.
IPR1	I know whether my institution has rules or guidance on generative AI use.

Source: supplied Group 4 workbook Codebook. Responses range from 1 strongly disagree to 5 strongly agree. GAI4 uses the Codebook wording “legal claims”, which differs slightly from the wording in the supplied article.

The analysis preserves the recorded responses and does not compute a validated overall responsibility, literacy, integrity or policy score.

Appendix B Threshold sensitivity

Table B1 reports the additional binary threshold summaries referred to in Section 3.6. The ordinal model remains the principal adjusted analysis.