Pedagogical Research and Development

Volume: 5 Issue: 4

2026

ISSN: 2791-3627



KMF PUBLISHERS

http://kmf-publishers.com/prd/

DOI: https://doi.org/10.64907/xkmf.v5i4.prd.1

RESEARCH ARTICLE

OPEN ACCESS
Freely available online

Generative AI Use and Academic Integrity Behaviour among Law Students in Bangladesh

Parvas Mia[1*]; Mst. Sadia Rahman Arpita[1]; Sifa Sultana[1]; Kazi Abdul Mannan[2]

Abstract: Generative artificial intelligence presents legal education with two distinct integrity tasks: checking the reliability of generated legal material and making assistance visible when disclosure is required. This study examines their relationship through an exploratory analysis of the Group 4 Law-SDG database. Of 385 records, 245 identify the respondent as a law student and constitute the primary sample. Four conceptually matched items distinguish verification from disclosure and stated responsibility from reported practice. Analyses combine exact paired tests, Holm adjustment, respondent-level bootstrap intervals and population-averaged logistic models with respondent-clustered covariance estimation. Verification practice is endorsed by 204 students (83.3%), compared with 160 (65.3%) for disclosure practice: a paired difference of 18.0 percentage points (95% bootstrap interval 10.2-25.7). The corresponding normative difference is 15.5 points (8.2-22.9). Within each domain, duty-practice differences are small and statistically uncertain. Adjusted endorsement odds are higher for verification than disclosure (OR 2.67, 95% CI 1.71-4.18), while policy awareness, training and use frequency have uncertain conditional associations. Only 43 students (17.6%) endorse all five integrity-practice items. Results persist with stricter endorsement and in wider-sample checks. Low block internal consistency supports item-level interpretation. Numerical provenance remains unverified because the codebook labels identifiers and contextual comments synthetic; the latter are excluded. The study identifies a verification-disclosure asymmetry in recorded responses, without estimating misconduct prevalence or testing the full theory of planned behaviour. It proposes separate educational workflows for checking legal authority and documenting permissible AI assistance.

Keywords: legal education; generative artificial intelligence; academic integrity; disclosure; verification; planned behaviour; Bangladesh

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

Corresponding author:
Parvas Mia
mdparbes777777@gmail.com

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

Citation information: Mia, P., Arpita, M. S. R., Sultana, S., & Mannan, K. A. (2026). Generative AI Use and Academic Integrity Behaviour among Law Students in Bangladesh. *Pedagogical Research and Development, 5*(4). https://doi.org/10.64907/xkmf.v5i4.prd.1

1. Introduction

Generative artificial intelligence can assist students with outlining, explanation, translation and drafting, but the educational value of these functions depends on the task and the student's continuing intellectual involvement. Higher education therefore faces a practical question about what learners should do before submitting AI-assisted work. General discussions of academic integrity often combine different obligations: obtaining permission, maintaining independent judgement, acknowledging assistance, checking information and accepting responsibility. Those obligations can support one another without being interchangeable. Research on AI-assisted assessment has accordingly moved towards explicit permissions, transparent expectations and assessment design rather than treating every use of a language model as misconduct.[1]

Legal education makes this distinction particularly consequential. A plausible paragraph may contain an invented authority, a misdescribed holding or a rule from the wrong jurisdiction. Verification is therefore an epistemic practice: it concerns whether a proposition is supported by an identifiable and applicable legal source. Disclosure is an attribution practice: it concerns whether the evaluator can understand the role of an external tool in producing the submission. Evidence of legal hallucinations establishes a reason to teach verification, while it does not by itself answer when disclosure is necessary or how a declaration should be made.[2]

The integrity literature also cautions against reducing behaviour to individual character. Institutional expectations, peer environments and the interpretation of rules matter alongside personal commitments. Earlier research on academic dishonesty documented contextual influences well before generative AI became widely accessible. The International Center for Academic Integrity provides a complementary values framework, including honesty, responsibility and fairness. These perspectives suggest that a student's response to one integrity obligation need not reveal their response to another. A learner might check citations while remaining uncertain about whether an editing interaction warrants acknowledgement. Another might provide a declaration while inadequately checking the legal content.[3]

[1] Enkelejda Kasneci and others, ‘ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education’ (2023) 103 Learning and Individual Differences 102274; Debby RE Cotton, Peter A Cotton and J Reuben Shipway, ‘Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT’ (2024) 61 Innovations in Education and Teaching International 228; 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, DOI: 10.53761/1.20.02.07.

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

[3] Donald L McCabe and Linda Klebe Treviño, ‘Academic Dishonesty: Honor Codes and Other Contextual Influences’ (1993) 64 Journal of Higher Education 522; Donald L McCabe, Linda Klebe Treviño and Kenneth D Butterfield, ‘Cheating in Academic Institutions: A Decade of Research’ (2001) 11 Ethics & Behavior 219; International Center for Academic Integrity, The Fundamental Values of Academic Integrity (3rd edn, ICAI 2021).

Recent student research reinforces the importance of specifying the practice being discussed. Johnston and colleagues examined student perspectives to inform institutional integrity guidance; their work distinguishes support for assistance from acceptance of wholesale substitution. Gruenhagen and colleagues likewise show why students' use of chatbots for assessment cannot simply be equated with a uniform understanding of misconduct. These studies justify asking how respondents interpret particular activities, rather than assuming that agreement with a broad statement establishes a stable integrity disposition.[4]

Bangladesh is an informative context for this question, but the contextual contribution must be stated precisely. Hasib and Islam's qualitative study describes varied motivations for ChatGPT use among university students and highlights concerns about references and responsible practice. It does not provide a national estimate for law students, nor does it establish the distribution of verification and disclosure behaviours examined here. The present study addresses that narrower measurement question using the Group 4 questionnaire and database prepared for a law-and-sustainable-development research project.[5]

The research plan names Parvas Mia's topic as generative AI use and academic integrity behaviour among law students. Its proposed theoretical foundation combines the theory of planned behaviour with an academic integrity framework. The common questionnaire contains four blocks, but the five integrity items describe distinct actions rather than repeated measures of one narrowly defined behaviour. The database also includes law graduates, legal professionals and other respondents. A valid analysis therefore requires both a population decision and a measurement decision: the primary sample should reflect the named student population, and the statistical model should not assume that heterogeneous obligations form a validated latent scale.

The central question is whether the recorded responses indicate a broad divide between responsibility and practice, or a more specific difference between verification and disclosure. This distinction changes the interpretation of the same favourable average. If all duties exceed reported practices, an implementation account becomes plausible. If verification exceeds disclosure in both duties and practices, the more immediate issue concerns the organisation and interpretation of different obligations. Neither pattern alone establishes its cause, but separating them prevents an aggregate score from supplying an explanation that the items do not support.

The study asks four questions. RQ1: How do law-student records distribute across the five integrity-practice items? RQ2: Is verification more strongly endorsed than disclosure within the same respondent, for both

[4] Heather Johnston and others, ‘Student Perspectives on the Use of Generative Artificial Intelligence Technologies in Higher Education’ (2024) 20 International Journal for Educational Integrity 2; Jan Henrik Gruenhagen and others, ‘The Rapid Rise of Generative AI and Its Implications for Academic Integrity: Students’ Perceptions and Use of Chatbots for Assistance with Assessments’ (2024) 7 Computers and Education: Artificial Intelligence 100273.

[5] Mir Hasib and Md Shariful Islam, ‘How University Students in Bangladesh Engage with ChatGPT: A Qualitative Study’ (2025) 20 PLOS ONE e0333089.

responsibility statements and reported-practice statements? RQ3: Are responsibility-practice contrasts different across these domains? RQ4: Does the verification-disclosure association persist after adjustment for measured policy awareness, AI awareness, training and use frequency, and under alternative analytical definitions?

The contribution is a paired, behaviour-specific account of integrity responses. It differs from adoption-centred models, studies primarily concerned with authorship responsibility, and institutional-policy analyses by keeping the outcome at the level of specific student practices. It also avoids treating the 385-record database as 385 law students. Its findings support differentiated teaching and reporting workflows, while the documented provenance uncertainty limits empirical publication claims. The article develops the theoretical distinction, reports the analytical design and results, and then considers educational implications and the evidence required for further validation.

2. Literature Review and Theoretical Foundation

This section connects planned-behaviour reasoning with academic integrity scholarship to distinguish the object of an obligation from the form in which a questionnaire asks about it. The aim is an interpretable comparison of verification and disclosure, rather than a comprehensive latent-variable test. The framework supplies theoretically motivated contrasts; the available items determine which contrasts can actually be estimated.

2.1 Integrity as a collection of practices

Academic integrity encompasses honesty about the origin of work, responsibility for its content and fairness in assessment. The ICAI framework treats these as educational values with practical implications rather than as a list of prohibited technologies. For AI-assisted legal assignments, these values can be operationalised through several decisions: whether a tool is permitted, what intellectual work must remain the student's, whether assistance should be declared, and how sources will be checked. A single approval question cannot resolve all these decisions.[6]

The distinction between verification and disclosure is useful because each addresses a different evidentiary problem. Verification asks whether a legal claim survives examination against authoritative material. Disclosure asks whether the representation of the student's contribution is sufficiently transparent under the applicable assessment

[6] ICAI (n 3).

instructions. A citation can be accurate even when the drafting process is undisclosed, and a transparent declaration can accompany inaccurate legal analysis. These are conceptual possibilities, not observed classifications of misconduct in this dataset. Their value lies in explaining why one practice should not be used as a proxy for the other.

AI literacy provides additional conceptual context. Long and Magerko describe competencies for critically evaluating and interacting with AI. Such competence is broader than use frequency and cannot be inferred from the possession of an account or familiarity with a product name. The Group 4 awareness items capture particular self-reported orientations, such as awareness of inaccuracy and bias. They do not objectively test legal research ability, and their inclusion as covariates should not be interpreted as a validated measure of comprehensive AI literacy.[7]

2.2 Planned behaviour and the specificity of the outcome

Ajzen's theory of planned behaviour distinguishes attitudes, subjective norms, perceived behavioural control and intentions in explaining a specified behaviour. Its relevance here is the requirement to identify what action is being considered. Checking a legal authority and declaring AI assistance are different actions with potentially different consequences, opportunities and social meanings. Beck and Ajzen's application to dishonest actions further demonstrates why integrity-related behaviour can be examined within a planned-behaviour perspective. However, theoretical relevance does not mean that every integrity questionnaire contains the constructs needed to test that theory.[8]

The present instrument does not provide separate validated measures of attitudes towards each action, perceived peer expectations, control over performing it, and temporally prior intentions. Consequently, the study does not claim a full theory-of-planned-behaviour test. The AR responsibility statements are not relabelled subjective norms: an individual's endorsement of a duty is different from perceived pressure from relevant others. Likewise, policy awareness is not equated with perceived behavioural control, and training attendance is not treated as a direct measure of capability. Maintaining these distinctions avoids creating theoretical constructs through convenient variable names.

[7] 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).

[8] Icek Ajzen, ‘The Theory of Planned Behavior’ (1991) 50 Organizational Behavior and Human Decision Processes 179; Lisa Beck and Icek Ajzen, ‘Predicting Dishonest Actions Using the Theory of Planned Behavior’ (1991) 25 Journal of Research in Personality 285.

Planned-behaviour reasoning nevertheless supports a useful interpretive proposition. Different integrity actions should be measured separately before their antecedents or aggregate consequences are modelled. The same learner may see verification as a normal research step and declaration as a communication decision with uncertain assessment consequences. These possible evaluations motivate comparison across actions, while the survey does not establish that either mechanism produced the responses. The resulting framework is deliberately narrower than a structural model of attitudes, intention and subsequent behaviour.

2.3 Institutional context and academic integrity

Research predating generative AI links academic dishonesty to contextual as well as individual factors. Whitley's review and McCabe and colleagues' work provide a basis for considering institutional environments, expectations and social influences. They do not establish that a particular rule or training programme will produce the same effect in every setting. Their relevance is to resist an exclusively dispositional interpretation of a student's response and to motivate collecting information about institutional conditions.[9]

The more recent declaration literature makes this issue concrete. Gonsalves identifies ambiguity, perceived consequences and inconsistency as barriers in a particular institutional setting. These findings suggest mechanisms that may be relevant to disclosure, but the current database does not contain interviews about fear, peer competition or institutional trust. They therefore inform explanatory possibilities rather than coding categories or measured mediators. The study's contribution is to ask whether a declaration-specific issue is visible in the distribution of responses before attributing it to any such barrier.[10]

Practical guidance also increasingly locates integrity within task design. ENAI recommendations emphasise responsible use and educational policies. The AI Assessment Scale offers a framework for making the permissible contribution of AI explicit, and its later refinement places greater emphasis on assessment redesign and communication. These approaches make an important conceptual point for the present study: compliance depends partly on the rule attached to a task, so a statement qualified by “when required” should not be interpreted independently of that conditional wording.[11]

[9] Bernard E Whitley Jr, ‘Factors Associated with Cheating among College Students: A Review’ (1998) 39 Research in Higher Education 235; McCabe, Treviño and Butterfield (n 3).

[10] Chahna Gonsalves, ‘Addressing Student Non-Compliance in AI Use Declarations: Implications for Academic Integrity and Assessment in Higher Education’ (2025) 50 Assessment & Evaluation in Higher Education 592.

[11] Tomas Foltynek and others, ‘ENAI Recommendations on the Ethical Use of Artificial Intelligence in Education’ (2023) 19 International Journal for Educational Integrity 12; 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; 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.

2.4 Legal verification and transparent assistance

Legal verification requires more than recognising a familiar case name. It involves locating the actual authority, checking the proposition attributed to it and considering relevance to the jurisdiction and problem. Dahl and colleagues' work on legal hallucinations establishes a concrete reason for distinguishing apparent plausibility from legal accuracy. The present GAI4 item asks whether respondents verify generated citations and claims; it does not observe which sources they consult or whether the verification succeeds. An endorsement therefore represents a reported orientation towards checking, not demonstrated legal competence.[12]

Disclosure presents a different assessment challenge. A student may need to communicate whether AI was used for brainstorming, language revision, drafting, or analysis, and whether generated text was retained. Academic integrity discussions by Perkins and Cotton and colleagues motivate clarifying these boundaries. Their work also supports avoiding a blanket assumption that assistance and substitution are identical. The Group 4 disclosure item is explicitly conditional on a requirement; a lower agreement score could reflect practice, uncertainty, a perceived absence of a requirement, or item interpretation. The survey cannot isolate these possibilities.[13]

Student participation in designing guidance is relevant because the meaning of a declaration must be comprehensible to its audience. Johnston and colleagues provide an example of using student perspectives to inform policy, while the AIAS pilot describes implementation in a specific educational setting. Neither source supplies an intervention effect for Bangladeshi law students. Together, they justify evaluating whether rules can be explained and used in actual assignments, rather than assuming that publishing a general statement constitutes implementation.[14]

Reliance on automated detection offers no substitute for this clarity. Weber-Wulff and colleagues evaluate limitations of detection tools, while Liang and colleagues identify fairness concerns for non-native English writing. These studies do not imply that every detector performs identically or that any particular student text has been misclassified. They do support caution about using detection outputs as definitive evidence of authorship. For legal education, process evidence and the ability to explain legal reasoning provide more direct objects of assessment than an unsupported inference from stylistic fluency.[15]

[12] Dahl and others (n 2).

[13] Perkins, ‘Academic Integrity Considerations’ (n 1); Cotton, Cotton and Shipway (n 1).

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

[15] Debora Weber-Wulff and others, ‘Testing of Detection Tools for AI-Generated Text’ (2023) 19 International Journal for Educational Integrity 26; Weixin Liang and others, ‘GPT Detectors Are Biased against Non-Native English Writers’ (2023) 4 Patterns 100779.

2.5 Analytical propositions and conceptual framework

The analysis develops two directional propositions and two exploratory questions. H1 proposes that reported verification practice receives greater endorsement than reported disclosure practice. This expectation is motivated by the distinctive legal importance of checking authorities and by literature identifying declaration-specific implementation problems. It is not a claim that disclosure has lesser ethical value. H2 proposes that the same ordering appears in responsibility statements. This examines whether the difference lies only in enactment or is also present in how duties are evaluated.

RQ3 compares responsibility and practice within each domain without presuming a general deficit. A normative duty statement and a self-report are not psychometrically identical, so a paired difference cannot be interpreted as a direct measure of hypocrisy or intention failure. The study instead uses their conceptual correspondence to locate differences in recorded endorsement. RQ4 examines adjusted domain associations alongside policy-rule awareness, awareness of AI inaccuracy and bias, training and frequency of use. These analyses are exploratory, not preregistered confirmatory tests.

Figure 1 represents the design as a two-by-two comparison. Its columns distinguish duty statements from reported-practice statements; its rows distinguish disclosure from verification. Vertical contrasts address H1 and H2, while horizontal contrasts address RQ3. A second analysis uses the two reported-practice items as repeated outcomes and examines the domain coefficient conditional on individual covariates. There are no mediation paths, latent factors or implied causal arrows. This design makes a specific empirical distinction visible while preserving the limits of the available instrument.

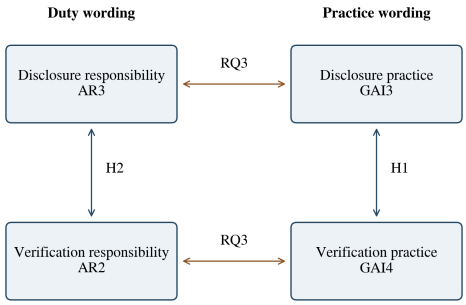

Figure 1. Paired conceptual design distinguishing domain from wording. H1 and H2 compare verification with disclosure; RQ3 compares duty and practice wording. Arrows denote contrasts, not causal pathways.

3. Methodology

The study is a secondary numerical analysis of a cross-sectional questionnaire database. It follows the supplied plan's substantive focus on law-student integrity while adapting the statistical strategy to the actual population labels and measurement properties. Reporting distinguishes documented database features from fieldwork procedures that cannot be independently established. This follows the general observational-reporting principle that the achieved sample and available evidence should be described explicitly.[16]

3.1 Source materials and analytical population

The analysis uses three project documents and the Group-4-Law-SDG-Database workbook: the research plan, the Q4 common questionnaire and the requested article structure. The workbook contains README, Data and Codebook sheets. Data comprise 385 records and 35 fields. The questionnaire supports three separate student topics; this article concentrates on the GAI practice block and the specific AR2/AR3 responsibility items needed to interpret verification and disclosure. Other authorship-responsibility items and institutional-policy outcomes are not treated as competing primary endpoints.

All 385 records contain affirmative consent, adult-eligibility and AI-awareness entries. Current status identifies 245 law students, 71 law graduates, 37 advocates or legal professionals, and 32 other respondents. The 245 self-identified law students constitute the primary population for the title. Education-level categories are retained descriptively rather than used to redefine current status. The remaining records enter only a wider-sample sensitivity analysis and do not increase the primary student sample size.

3.2 Sampling and evidentiary scope

The plan specifies Cochran's large-population proportion calculation with Z = 1.96, p = .50 and e = .05, yielding 384.16 and a rounded target of 385 completed responses. It anticipates approximately 428 contacts for 10% non-response. These are planning parameters, not an observed recruitment history. No documented selection probabilities, institutional frame, contact denominator, response rate or collection dates accompany the numerical file. The achieved student sample is 245, not the planned 385; no retrospective power claim is made.[17]

A further provenance issue concerns the Codebook descriptions “Synthetic respondent sequence” and “Synthetic contextual comment.” The identifier label could describe generated anonymisation, but the documentation does not independently establish the origin of the numerical responses. Optional comments are excluded completely; they are not analysed as participant testimony. Numerical estimates are reported as characteristics of the supplied records. Confirmation of original collection and provenance is necessary before

[16] 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.

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

representing the analysis as verified empirical evidence about human participants.

3.3 Measures and analytical coding

Responses use an agreement scale from 1, strongly disagree, to 5, strongly agree. GAI1 concerns permitted use, GAI2 independent-work boundaries, GAI3 required disclosure, GAI4 verification of citations and legal claims, and GAI5 avoiding presentation of substantially generated work as entirely one's own. These items define practices with different objects. Table 1 identifies the focal pairings and covariates; Appendix A reproduces the focal wording to make the interpretation auditable.

Endorsement is defined as 4 or 5, with 1-3 grouped as “not endorsing at this threshold.” Neutral responses are not classified as admissions of misconduct. AR2 and GAI4 form the verification correspondence; AR3 and GAI3 form the disclosure correspondence. AR items concern responsibility, while GAI items are phrased as personal practice. Because both use agreement responses, the article reports endorsement rather than verified behavioural prevalence. Strict endorsement, defined as 5 only, and an ordinal sign test assess sensitivity to dichotomisation.

Table 1. Focal measures and analytical roles

Code	Content	Analytical role
GAI1-GAI5	Five integrity practices	Separate descriptive outcomes
GAI3 / AR3	Disclosure practice / responsibility	Matched disclosure correspondence
GAI4 / AR2	Verification practice / responsibility	Matched verification correspondence
IPR1	Knowledge of institutional AI rules	Observed-item covariate
AIL1 / AIL3	Awareness of inaccuracy / bias	Separate observed-item covariates
D8 training	Any orientation or course versus none	Binary exposure covariate
D7 frequency	Never to daily or almost daily	Ordered exposure covariate

Source: Q4 common questionnaire. The questionnaire supplies the item wording; theoretical references inform interpretation rather than establish that these items are validated adaptations. No item is reverse-coded.

3.4 Screening and measurement diagnostics

Screening examines identifier duplication, exact duplication of the 20 substantive responses, missing values, permitted ranges and constant responding across all 20 items. These checks do not authenticate field collection or prove that each record represents a different person. There is no dedicated attention-check item; F2 asks about confidence in understanding and is not treated as an attention test. No exclusion is based on an invented completion-time threshold.

Cronbach's alpha is calculated separately for AIL, GAI, AR and IPR in both samples. Alpha is a diagnostic of the responses to an item set, not proof of validity or a direct test of unidimensionality. The analysis does not delete items to maximise alpha. Given heterogeneous content and low internal consistency, it does not estimate reflective SEM, composite reliability, AVE or HTMT as though a validated measurement model had been established. The focal outcomes remain individual items.[18]

3.5 Paired tests and repeated-response models

Four paired contrasts are examined: verification minus disclosure within reported practice; the same contrast within responsibility; responsibility minus practice for verification; and responsibility minus practice for disclosure. Exact two-sided McNemar tests condition on discordant pairs. Holm's procedure adjusts the four student-sample p values as one family. Percentage-point differences receive percentile intervals from 20,000 bootstrap samples of complete respondents, using seed 20261019. Intervals are pointwise, whereas multiplicity adjustment applies to the tests.[19]

A population-averaged logistic model stacks the four focal responses per student, producing 980 observations within 245 respondent clusters. Predictors are domain, wording and their interaction. Verification is coded 1, disclosure 0; reported practice is 1, duty wording 0. Logistic mean equations use working independence, with a respondent-clustered sandwich covariance and the finite-cluster multiplier G/(G-1). This is a working-independence GEE implementation; it does not assume independence for variance estimation and does not represent longitudinal observations.[20]

A second model stacks the two practice items, producing 490 responses. It includes domain, IPR1, AIL1, AIL3, any training and frequency of use. Item covariates are centred at 3; frequency runs from 0, never, to 4, daily or almost daily, centred at 2. Any orientation or course counts as training. AIL2 is excluded because it directly asks about verification, which overlaps the outcome. No interaction terms or data-driven covariate selection are used in this adjusted model. Numerical optimisation uses the logistic score and checks convergence against near-zero score residuals.

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

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

[20] Kung-Yee Liang and Scott L Zeger, ‘Longitudinal Data Analysis Using Generalized Linear Models’ (1986) 73 Biometrika 13; A Colin Cameron and Douglas L Miller, ‘A Practitioner’s Guide to Cluster-Robust Inference’ (2015) 50 Journal of Human Resources 317.

3.6 Sensitivity, interpretation and ethics

Sensitivity checks use the complete 385-record sample, strict endorsement, an ordinal sign test, and an adjusted model excluding nine students reporting never using generative AI. None is an independent replication. Odds ratios describe endorsement odds, not risk ratios or causal intervention effects. Institution-level clustering cannot be estimated because institutional identifiers are unavailable. Covariate analyses are exploratory; uncertainty is assessed through estimates and intervals rather than treating a p-value threshold as a measure of importance.[21]

This secondary analysis involved no new recruitment. The questionnaire specifies voluntary participation, anonymous responses and the option to stop. Affirmative screening fields are documented, but original consent records, formal ethics approval or exemption, and fieldwork implementation are not available for independent verification. No formal clearance is inferred from the study's educational purpose. The analysis excludes optional comments, avoids identity inference and reports aggregates. These protections govern the present use of the file without claiming that they retrospectively validate undocumented original collection.

4. Results

Results first define the sample and measurement scope, then report the four paired contrasts and repeated-response models. The student-only sample is primary throughout. Tables distinguish raw endorsement counts from model-based odds ratios, and sensitivity results retain their separate denominators. The interpretation concerns the recorded responses; it does not resolve the provenance limitations described above.

4.1 Sample profile and data checks

Law students account for 245 of 385 records (63.6%). The largest student age category is 21-23 years, with 102 records (41.6%), followed by 24-26 years with 57 (23.3%). There are 131 male and 110 female student records, with two in each remaining gender category. Dhaka accounts for 92 students (37.6%). Undergraduate status is recorded for 139, while the other educational categories are graduate, postgraduate or professional. These categories show that self-identification as a current law student is not identical to undergraduate classification.

Table 2 summarises the student profile and exposure variables. Weekly or daily AI use is recorded for 171 students (69.8%); nine (3.7%) report never using it. No formal training is reported by 129 students (52.7%), a short orientation by 78 (31.8%), and a course or certified training by 38 (15.5%). These are exposure descriptors rather than validated measures of competence or institution-specific policy access.

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

Table 2. Profile of the primary law-student sample (n = 245)

Profile	Category order	Count in category order
Age	18-20 / 21-23 / 24-26	37 / 102 / 57
Age	27-30 / 31+	24 / 25
Gender	Male / female / other / prefer not to say	131 / 110 / 2 / 2
Education	Undergraduate / graduate	139 / 62
Education	Postgraduate / professional	15 / 29
Location	Dhaka / other city corporation / district town	92 / 47 / 64
Location	Upazila-rural / other	36 / 6
Use frequency	Never / less than monthly / monthly	9 / 24 / 41
Use frequency	Weekly / daily or almost daily	85 / 86
Training	None / short orientation / course or certified	129 / 78 / 38

No substantive response is missing or outside 1-5. The file contains 385 unique identifiers, no exact duplicate 20-item vectors and no record with a constant response across all 20 substantive items. Fourteen missing values occur in the AI-tool field, which is not used in the models. These results permit complete-case analysis of the defined variables without imputation. They establish coding completeness for those variables, not participant authenticity or probability recruitment.

4.2 Item patterns and measurement boundaries

Student means range from 3.727 to 4.167 across the GAI items. Verification has the highest practice mean and endorsement rate. Permitted use has the lowest rate at 61.2%; disclosure and independent-work boundaries each receive 65.3% endorsement, while avoiding entirely-own attribution for substantially generated work receives 68.6%. Thus the practice distribution is differentiated even though every mean exceeds the scale midpoint. Table 3 reports focal response summaries alongside awareness covariates.

Table 3. Focal item descriptions and student responses

Item and abbreviated content	Mean	SD	Endorsement n (%)
GAI1 Permitted use	3.727	0.796	150 (61.2%)
GAI2 Independent-work boundaries	3.820	0.815	160 (65.3%)
GAI3 Required disclosure practice	3.796	0.814	160 (65.3%)
GAI4 Verification practice	4.167	0.758	204 (83.3%)
GAI5 Avoid entirely-own attribution	3.812	0.818	168 (68.6%)
AR2 Verification responsibility	4.192	0.735	202 (82.4%)
AR3 Disclosure responsibility	3.816	0.781	164 (66.9%)
IPR1 Policy-rule awareness	3.714	0.820	151 (61.6%)
AIL1 Inaccuracy awareness	3.673	0.809	147 (60.0%)
AIL3 Bias awareness	4.208	0.720	208 (84.9%)

All-five-item endorsement occurs in 43 students (17.6%). This conjunctive profile is reported as a response configuration, not a validated integrity score or diagnostic classification. In particular, not endorsing all five items does not establish that the

remaining 202 students committed misconduct. The result demonstrates that favourable responses to individual practices do not automatically imply endorsement of the entire bundle.

Student-sample alpha values are .218 for AIL, .198 for GAI, .309 for AR and .162 for IPR. Corresponding full-sample values are .205, .301, .225 and .135. These values constrain interpretation of block averages as coherent reflective traits. They do not identify whether item breadth, wording, measurement error or other characteristics account for the pattern. The analysis consequently proceeds with the specified item comparisons rather than latent-factor paths.

4.3 Paired differences between domains

Verification practice is endorsed by 204 students and disclosure practice by 160. Of the 104 discordant pairs, 74 endorse verification alone and 30 disclosure alone. The difference is 18.0 percentage points, with a bootstrap interval of 10.2-25.7 points. The exact test yields p = .0000189 and Holm-adjusted p = .0000758. This supports H1 within the recorded student responses. There are 130 students endorsing both and 11 endorsing neither at the threshold.

Responsibility statements exhibit the same ordering. AR2 verification responsibility is endorsed by 202 students, compared with 164 for AR3 disclosure responsibility. There are 65 verification-only and 27 disclosure-only pairs. The paired difference is 15.5 points (8.2-22.9); exact p = .0000926 and Holm-adjusted p = .000278. H2 therefore receives support. Table 4 presents both domain contrasts and the within-domain wording contrasts with a common interpretation of the signs.

Table 4. Primary paired contrasts with multiplicity adjustment

Contrast (first - second)	First only / second only	Gap (pp)	95% bootstrap interval	Holm p
Practice: verification - disclosure	74 / 30	18.0	10.2-25.7	< .001
Duty: verification - disclosure	65 / 27	15.5	8.2-22.9	< .001
Verification: duty - practice	32 / 34	-0.8	-7.3-5.7	1.000
Disclosure: duty - practice	54 / 50	1.6	-6.5-9.8	1.000

Note: pp = percentage points. Positive gaps indicate greater endorsement of the first item. Intervals are pointwise percentile bootstrap intervals; Holm adjustment applies to the family of four exact paired tests.

The within-domain comparisons do not show a corresponding general excess of responsibility endorsement over practice. Verification responsibility is 0.8 points lower than practice, with an interval from -7.3 to 5.7 points. Disclosure responsibility is 1.6 points higher, with an interval from -6.5 to 9.8. Holm-adjusted p values equal 1.000 for both comparisons. These intervals are compatible with differences in either direction; the results are not equivalence tests.

Figure 2 displays the four contrasts on the same percentage-point axis. Its main pattern is that the two domain contrasts are positive, whereas the wording contrasts are centred close to zero with intervals crossing zero. This graphical comparison separates a practice-specific asymmetry from a generic interpretation that respondents endorse duties they do not enact. The ordinal and

stricter-threshold analyses below assess whether this ordering depends on the main endorsement definition.

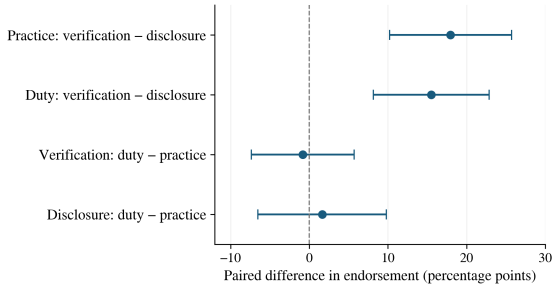

Figure 2. Student-sample paired endorsement differences and 95% pointwise bootstrap intervals. The vertical line denotes no difference; domains and wording contrasts are displayed separately.

4.4 Repeated-response model

The four-response model estimates higher endorsement odds for the verification domain within duty wording: OR 2.32 (95% CI 1.53-3.52, p < .001). The reported-practice wording coefficient, evaluated in the disclosure domain, is OR 0.93 (0.65-1.34, p = .695). The domain-by-wording interaction is OR 1.14 (0.62-2.10, p = .675). Table 5 reports these terms with their reference categories. The fitted probabilities reproduce the four observed endorsement proportions, providing an internal check on the saturated two-by-two mean specification.

Table 5. Four-response population-averaged logistic model

Term	OR	95% CI	p
Verification domain	2.32	1.53-3.52	< .001
Reported-practice wording	0.93	0.65-1.34	0.695
Domain × wording	1.14	0.62-2.10	0.675

Reference cell: disclosure responsibility. Domain OR refers to verification versus disclosure within duty wording; practice OR refers to practice versus duty within disclosure. Interaction is a ratio of odds ratios. Model uses 980 responses in 245 respondent clusters.

The interaction interval includes appreciable differences in either direction. It therefore provides limited evidence that the verification-disclosure odds contrast differs by wording, without proving that the contrasts are identical. This is consistent with the paired comparisons, which locate the clearer separation between domains. Respondent clustering ensures that four answers from one person do not contribute four independent variance units.

4.5 Adjusted associations and sensitivity checks

The adjusted two-practice model estimates a verification-domain OR of 2.67 (1.71-4.18, p < .001). IPR1 policy-rule awareness has OR 1.21 per agreement point (0.96-1.53, p = .104). The estimates for awareness of inaccuracy, bias awareness, any formal training and use-frequency step all have intervals crossing 1. Table 6 presents the full specification, including these uncertain results. No coefficient is interpreted as the effect of providing a policy or training intervention.

Table 6. Adjusted repeated-practice model among law students

Predictor	OR	95% CI	p
Verification domain	2.67	1.71-4.18	< .001
Policy-rule awareness IPR1	1.21	0.96-1.53	0.104
Inaccuracy awareness AIL1	0.94	0.73-1.20	0.602
Bias awareness AIL3	1.11	0.82-1.48	0.504
Any formal training	0.83	0.56-1.23	0.354
AI-use frequency step	1.12	0.93-1.35	0.229

Model uses 490 responses in 245 respondent clusters. Item covariates enter per one-point increment; frequency enters per one-category increment. Training compares any orientation/course with none. All intervals use respondent-clustered sandwich covariance; covariate analyses are exploratory.

Table 7 summarises robustness checks. In the complete 385-record sample, the practice-domain gap is 20.3 points (14.3-26.2), and the normative-domain gap is 18.4 points (12.7-24.4). Under strict student endorsement, the practice gap is 16.3 points (8.6-24.1). Retaining the ordinal ordering, 113 students rate verification above disclosure, 55 below and 77 equally; the exact sign-test p value is .00000902. Among the 236 students reporting some AI use, the adjusted domain OR is 2.59 (1.65-4.09).

Table 7. Sensitivity analyses and joint-response profile

Analysis	n	Estimate or counts	95% interval / detail
All records: practice gap	385	20.3 pp	14.3-26.2
All records: duty gap	385	18.4 pp	12.7-24.4
Student strict endorsement: practice gap	245	16.3 pp	8.6-24.1
Student ordinal sign test	245	113 higher / 55 lower	77 tied; p < .001
Student AI users: adjusted domain OR	236	2.59	1.65-4.09
All-five-practice endorsement	245	43 students	17.6%

Gap intervals are respondent-level bootstrap intervals. Wider-sample analyses include non-students and are not substituted for the primary analysis. The ordinal test uses the 168 non-tied student pairs. The all-five profile is descriptive, without a validated integrity threshold.

The combined results show a recurring verification-disclosure ordering, alongside uncertain responsibility-practice differences and uncertain covariate associations. They do not establish a population prevalence of non-compliance or explain why particular students responded differently. The wider-sample analysis supports consistency of the recorded ordering, while the student-restricted design remains necessary for the article's substantive population claim.

5. Discussion

The study identifies a specific asymmetry rather than a uniform integrity deficit. Verification of AI-generated legal material receives stronger endorsement than disclosure in both reported practice and responsibility statements. The similarity of this ordering across wording forms matters because it limits a simple account in which students know every duty but fail to act. The data are more consistent with differentiated responses to particular obligations, although item wording, unmeasured conditions and provenance uncertainty prevent a definitive behavioural explanation.

5.1 Integrity depends on which action is specified

The 18.0-point practice difference is substantial within the recorded sample and survives several alternative analyses. Its interpretation is not that verification is sufficient for integrity or that disclosure is optional. It is that an overall integrity judgement would conceal unequal endorsement of two relevant actions. The five-item GAI mean is approximately 3.864, but such an average cannot reveal that verification stands at 83.3% endorsement while several other practices are closer to two-thirds. This finding illustrates the analytical value of retaining obligations at the item level.

The result connects with the behaviour-specific orientation of planned-behaviour theory. If the action differs, the relevant beliefs and opportunities may differ as well. Here, the framework directs attention to verification and declaration as separate targets for explanation. It does not allow responsibility items to substitute for the theory's distinct social-norm and control constructs. Future theoretical tests should therefore measure intention towards a particular declaration or checking task and examine subsequent performance of that same task.[22]

This contribution is narrower than establishing a new integrity theory, but useful for empirical design. An explanatory model can be technically complex while leaving its outcome conceptually unclear. Separating actions before modelling their antecedents makes it possible to ask whether an intervention changes verification, declaration, both, or neither. The present repeated-item analysis supplies the descriptive distinction needed for that later work; it does not supply the missing longitudinal sequence.

5.2 Verification as an epistemic responsibility in legal work

Verification has a clear disciplinary object: the legal authority and the proposition for which it is used. A learner can in principle compare an AI-generated statement with an actual judgment, statute or reliable commentary. This offers a plausible reason why checking may be more readily recognised as a duty than documenting the role of a drafting assistant. The database does not test that reason, and a high agreement response does not establish that respondents perform an adequate jurisdictional or citation check.

[22] Ajzen (n 8).

The legal-hallucination literature makes the educational relevance of verification concrete. It supports distinguishing the fluency of generated text from the accuracy of its legal claims. The present item asks about both citations and claims, so its endorsement should not be narrowed to checking that a reference exists. A sound teaching response would evaluate the proposition, the authority and the fit to the problem. That response is justified by the combined conceptual and external evidence, rather than by assuming that the survey observed successful verification.[23]

The distinction also protects against an overly favourable interpretation. Students may recognise that verification is desirable while using superficial checks, or may endorse a statement they understand normatively rather than behaviourally. The five-point agreement format cannot distinguish those possibilities. Reporting the result as verification endorsement keeps the finding informative without converting it into evidence that 83.3% of students reliably validate all AI-assisted legal work.

5.3 Disclosure as a separate communication task

Disclosure requires the student to interpret both the assistance and the relevant instruction. A declaration may describe an interaction rather than identify a quoted source; the appropriate level of detail depends on what the assessment permits and asks the student to demonstrate. This makes a declaration form a communication mechanism, not merely an extra citation field. It should enable an evaluator to understand how assistance relates to the learning outcome and the student's responsibility for the final work.

Gonsalves' study provides contextual evidence that declaration processes can be affected by ambiguity and perceived consequences. The current results are compatible with such mechanisms but do not identify them in Bangladesh. There is no measured fear variable, observation of instruction quality or record of actual submission declarations. The parallel ordering in responsibility statements additionally suggests caution about attributing the entire practice difference to an intention-to-action failure. Respondents may interpret the disclosure duty itself more conditionally than the verification duty.[24]

The wording “when required” is central. A student who believes no declaration was required may not endorse the practice item strongly even while behaving consistently with their understanding of the task. Conversely, a high response may express an intention to comply without documenting

[23] Dahl and others (n 2).
[24] Gonsalves (n 10).

any actual disclosure. Institutional follow-up should therefore link a specific assignment's permission and declaration requirements to student understanding before classifying a response as non-compliance. That is an implication for measurement and implementation, not an inference that the present sample lacked rules.

5.4 Why the responsibility-practice null results matter

Neither responsibility-practice comparison supports a general normative excess. This result prevents the discussion from presenting a familiar but untested story of students expressing ethical commitments while doing otherwise. It does not prove that commitments and actions coincide, because the items are not observed actions and the intervals permit meaningful differences. The correct interpretation is that these responses provide limited evidence for a within-domain wording gap at the chosen threshold. Non-significance is not evidence of equivalence.[25]

The interaction result adds a related limitation. Its wide interval means that the study cannot precisely quantify how differently the two domains respond to a switch in wording. The domain contrasts are estimated more clearly than the interaction. This pattern directs attention towards the finding the data support without concealing the uncertainty in the more elaborate model. It also illustrates why statistical sophistication should not be equated with a requirement that every model term be significant.

The conjunction result supplies a different descriptive lesson. Only 17.6% endorse every GAI practice, despite favourable means on each. A set of partial commitments does not necessarily constitute a complete process. Yet it would be inappropriate to interpret the complement as an 82.4% misconduct rate: the threshold includes neutral responses, item requirements are conditional, and no submitted work was inspected. The profile is useful for designing a comprehensive educational workflow, not for identifying offenders.

5.5 Policy awareness and training require cautious interpretation

The adjusted domain association persists when measured awareness and exposure variables enter the model. Their uncertain coefficients do not show that policy and training are ineffective. IPR1 records knowing whether rules exist, rather than assessing comprehension of a particular rule. Training categories do not describe curriculum, duration, attendance verification or skill acquisition. Frequency describes exposure rather than purpose, permitted use or quality of engagement. These variables are too coarse to evaluate interventions causally.

Policy-design frameworks such as AIAS and ENAI recommendations are therefore best used here as resources for developing testable interventions. A targeted legal assignment could state allowed assistance, require a proportionate declaration and assess an authority-checking record. Its evaluation would then compare actual

[25] Douglas G Altman and J Martin Bland, ‘Absence of Evidence Is Not Evidence of Absence’ (1995) 311 BMJ 485.

outputs and student understanding under documented conditions. The survey motivates differentiating those components, while it does not establish that adopting any named framework will reproduce the observed associations.[26]

5.6 Integrated interpretation and evidentiary limits

The analysis contributes a practice-specific map of the supplied records: verification is relatively strongly endorsed, disclosure less so, and duties are not uniformly endorsed more strongly than reported practices. The same ordering survives stricter endorsement, an ordinal test and the wider sample. These checks reduce concern that one binary cut-point or the inclusion of non-students alone produced the finding. They cannot correct undocumented selection, establish item validity or authenticate collection.

Bangladesh-specific qualitative research indicates why varied student motivations and academic tasks deserve attention. The present study adds a different analytical question rather than claiming to reproduce that qualitative evidence. Together, such approaches motivate combining interviews about requirements with observation of actual assignments and verification tasks. The corrected publication record for the contextual study is acknowledged in the bibliography.[27]

Most fundamentally, numerical precision and provenance are separate. A reproducible calculation can accurately describe an undocumented file. Because the codebook uses synthetic labels for identifiers and comments without independently documenting numerical origin, the results should remain an exploratory database analysis until that ambiguity is resolved. This limitation is not a statistical finding that the numerical data are fabricated; it is a boundary on what the available documentation establishes.

6. Theoretical and Practical Implications

The implications follow the level of the evidence: the study distinguishes response patterns across integrity actions and proposes ways to measure and teach those actions more clearly. It does not establish causes of misconduct or validate an institutional intervention. The practical recommendations are therefore framed as designs for implementation and evaluation rather than as effects already demonstrated in this sample.

[26] Perkins and others, ‘AI Assessment Scale’ (n 11); Foltynek and others (n 11).

[27] Hasib and Islam, ‘How University Students’ (n 5); Mir Hasib and Md Shariful Islam, ‘Correction: How University Students in Bangladesh Engage with ChatGPT: A Qualitative Study’ (2026) 21 PLOS ONE e0344589.

6.1 Theoretical implications

First, academic integrity models should distinguish the content of the obligation from the grammatical form of the question. The two-by-two design reveals whether differences follow verification versus disclosure, or responsibility versus practice. This separation offers a way to prevent constructs from absorbing several meanings before their relationships are estimated. Low internal consistency reinforces the need for this discipline, but item heterogeneity is also visible conceptually. Better measurement should begin with definitions, not merely attempts to improve a coefficient.

Second, planned-behaviour applications in AI-assisted education should match predictors and outcomes to a specific action. Declaration intention and perceived control over declaring assistance should predict a documented declaration task, while verification beliefs should be linked to an observed checking task. A general integrity score cannot be assumed to preserve those action-specific relationships. The present findings motivate that research design without claiming that its theoretical pathways have been tested.

Third, legal education adds a distinctive epistemic dimension to integrity. Accurately representing one's contribution and validating the authority of a legal proposition are different responsibilities. A model that accounts only for originality may miss verification failures; a model that accounts only for accuracy may miss hidden delegation. Treating both as explicit outcomes can connect academic integrity scholarship to the disciplinary standards of legal reasoning while avoiding a claim that legal students are inherently more or less ethical than other students.

6.2 Educational practice

Law teachers could introduce a two-part submission process. One part would ask students to identify AI functions used and the extent to which generated material entered the work. The other would ask them to document checks on selected authorities and explain the proposition each source supports. The study's domain distinction supplies the rationale for keeping these records separate. The format should remain proportionate to the assignment, avoiding burdensome documentation of trivial assistance that the instructions do not require.

A declaration should describe permissible activity without implying that disclosure authorises otherwise prohibited delegation. Likewise, an authority-check record should require reasoning rather than a list of links. An oral explanation of one selected claim could assess whether the student understands the source and its relevance. These are proposed assessment mechanisms, whose feasibility and effects should be piloted rather than assumed.

Training could use paired examples: accurate AI-assisted prose with inadequate attribution, and openly declared prose with a fabricated authority. Students would discuss why each raises a different issue. Feedback

should make the required remedy clear: improve the declaration in one case, verify or correct the legal content in the other. Such examples can make integrity operational without relying on automated detection as decisive proof. External research on detector limitations and linguistic fairness supports that caution.[28]

6.3 Institutional and societal implications

Institutions should communicate AI permissions at assignment level, provide accessible examples and ensure students can ask for clarification before submission. Student involvement in evaluating those instructions can identify ambiguous terminology. UNESCO's guidance and the AIAS literature supply broader resources for a human-centred and task-sensitive approach; they do not replace local curricular judgement.[29]

These recommendations align conceptually with quality education and accountable institutions under SDG 4 and SDG 16. No SDG indicator is measured, so the study does not estimate progress towards either goal. Its contribution is more specific: making the responsibilities attached to AI-assisted legal learning visible, assessable and distinguishable from one another. Evaluation should examine learning, transparency, equitable access and accuracy together, rather than relying on a lower detector score or a more favourable perception alone.

7. Conclusion Limitations and Future Research

The Group 4 database records a consistent ordering between verification and disclosure among 245 self-identified law students. Verification practice receives 83.3% endorsement and disclosure practice 65.3%, a paired difference of 18.0 percentage points. A similar difference appears between responsibility statements. In contrast, within-domain responsibility-practice differences are small and uncertain. The adjusted analysis retains a verification-domain association, and alternative endorsement and sample definitions preserve the ordering.

The central contribution is to distinguish which integrity action is endorsed from whether the questionnaire asks about a duty or personal practice. This reframing matters because the same data do not support a generic assertion that students endorse integrity principles they fail to enact. Nor do the data establish a rate of cheating. The study instead identifies a pattern in agreement responses and explains why attribution and legal accuracy require separate educational attention.

Several limitations define the conclusion's reach. Provenance of the numerical responses is unresolved in the supplied documentation. Recruitment probabilities, fieldwork dates, response rates and institutional clustering are unavailable. Primary-sample membership relies on current-status self-identification. All

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

[29] Fengchun Miao and Wayne Holmes, Guidance for Generative AI in Education and Research (UNESCO 2023); Perkins, Roe and Furze, ‘Reimagining’ (n 11).

outcomes are self-reports without inspection of assignments, AI interactions or actual declarations. The disclosure item is conditional on a requirement that is not separately documented. Low internal consistency limits block-level interpretation, while single-item outcomes leave wording and measurement error unresolved. Bootstrap and sandwich intervals do not incorporate these uncertainties.

Future work should first document original collection, consent and applicable review arrangements, then use a verified sampling strategy with institution identifiers. Cognitive interviews could establish how students understand verification, disclosure and the phrase “when required.” A revised instrument should separately assess attitudes, perceived norms, control and intentions for each action rather than retrospectively mapping broad items onto theory.

The next substantive test should use actual tasks. Students could receive assignments with explicit AI permissions and be asked to produce both a declaration and an authority-check record. Independent assessment could evaluate whether the declaration accurately describes assistance, whether the cited source exists, and whether it supports the legal proposition. A prospective design could examine whether clarification or training changes these outcomes without conflating greater disclosure with greater AI use.

Comparative research could examine undergraduate and postgraduate legal study, other disciplines, different linguistic settings and different institutional policies. Such comparisons require comparable instruments and sufficient sampling information, not merely separate group means. Until those steps are completed, the present article offers an auditable analytical framework and conditional database findings. Its practical direction is clear: teach and evaluate transparency about assistance alongside verification of legal content, while treating them as distinct responsibilities.

8. Declarations

Ethics and informed consent

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

Data availability

The source workbook and questionnaire are held within the research project. The numerical-data provenance requires clarification before public empirical use. Access to row-level data depends on verified ownership, consent and confidentiality conditions. Analytical definitions and supplementary specifications appear in the appendices.

Funding and competing interests

The supplied project documents do not record funding arrangements or declarations of competing interests; no absence of either is inferred.

Author contributions

Parvas Mia: Lead student researcher; conceptualisation; preparation of study materials; organisation and curation of survey data; manuscript development. The project was conducted under the supervision of Prof. Dr Kazi Abdul Mannan.

Kazi Abdul Mannan: Conceptual and methodological guidance; supervision; interpretation of statistical findings; critical review and editing of the manuscript.

Generative AI disclosure

Generative AI assisted literature organisation, analytical programming, manuscript drafting and language editing. Calculations were performed programmatically. Responsibility for verifying the underlying records, interpretation, references and final text remains with the authors.

Bibliography

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

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

Beck L and Ajzen I, 'Predicting Dishonest Actions Using the Theory of Planned Behavior' (1991) 25 Journal of Research in Personality 285

Cameron AC and Miller DL, 'A Practitioner's Guide to Cluster-Robust Inference' (2015) 50 Journal of Human Resources 317

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

Cotton DRE, Cotton PA and Shipway JR, 'Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT' (2024) 61 Innovations in Education and Teaching International 228

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

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

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

Foltynek T and others, 'ENAI Recommendations on the Ethical Use of Artificial Intelligence in Education' (2023) 19 International Journal for Educational Integrity 12

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

Gonsalves C, 'Addressing Student Non-Compliance in AI Use Declarations: Implications for Academic Integrity and Assessment in Higher Education' (2025) 50 Assessment & Evaluation in Higher Education 592

Gruenhagen JH and others, ‘The Rapid Rise of Generative AI and Its Implications for Academic Integrity: Students’ Perceptions and Use of Chatbots for Assistance with Assessments’ (2024) 7 Computers and Education: Artificial Intelligence 100273

Hasib M and Islam MS, ‘How University Students in Bangladesh Engage with ChatGPT: A Qualitative Study’ (2025) 20 PLOS ONE e0333089

, ‘Correction: How University Students in Bangladesh Engage with ChatGPT: A Qualitative Study’ (2026) 21 PLOS ONE e0344589

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

International Center for Academic Integrity, *The Fundamental Values of Academic Integrity* (3rd edn, ICAI 2021)

Johnston H and others, ‘Student Perspectives on the Use of Generative Artificial Intelligence Technologies in Higher Education’ (2024) 20 International Journal for Educational Integrity 2

Kasneci E and others, ‘ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education’ (2023) 103 Learning and Individual Differences 102274

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)

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

McCabe DL, Treviño LK and Butterfield KD, ‘Cheating in Academic Institutions: A Decade of Research’ (2001) 11 Ethics & Behavior 219

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, DOI: 10.53761/1.20.02.07

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

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

Whitley BE Jr, 'Factors Associated with Cheating among College Students: A Review' (1998) 39 Research in Higher Education 235

Appendix A Focal item wording and interpretation

Table A1. Exact wording of focal substantive items

Code	Questionnaire statement
GAI1	I use generative AI only in ways permitted by my course or institution.
GAI2	I distinguish between AI assistance and work that I must complete independently.
GAI3	I disclose or acknowledge AI assistance when required.
GAI4	I verify AI-generated citations and legal claims before using them in academic work.
GAI5	I avoid submitting substantially AI-generated work as entirely my own.
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.
IPR1	I know whether my institution has rules or guidance on generative AI use.
AIL1	I understand that generative AI can produce plausible but inaccurate information.
AIL3	I understand that AI outputs may reflect bias or incomplete data.

All items use the common five-point agreement scale. These statements assess self-report or endorsement; they are not observations of submitted work. D7 and D8 categories appear in Table 2.

Appendix B Reproducible analytical specification

Primary records satisfy D3_Current_Status = "Law student." Endorsement equals 1 when the response is 4 or 5 and 0 otherwise. Verification practice is GAI4; disclosure practice is GAI3; verification responsibility is AR2; disclosure responsibility is AR3. Strict endorsement equals 1 only at response 5. All focal variables are complete in both samples.

For respondent i and item j, the four-response mean model is logit Pr(Yij =

1) = β0 + β1 Verificationij + β2 Practiceij + β3 Verificationij × Practiceij. The covariance is G/(G-1) A■⁻¹(Σi UiUi′)A■⁻¹, where A = Σij pij(1-pij)xijxij′ and Ui = Σj xij(Yij-pij). The respondent, rather than the item response, defines the independent cluster. The adjusted practice model replaces the wording terms with the six coefficients listed in Table 6, including domain.

Exact paired tests use the two discordant cell counts as a conditional binomial experiment with probability .5. Holm adjustment covers four primary contrasts. Bootstrap sampling resamples complete respondents 20,000 times with seed 20261019; the 2.5th and 97.5th percentiles form pointwise intervals. Model score residual maxima are below 0.000003 in the reported fits. In the saturated four-cell model, fitted proportions match observed proportions to numerical precision.

The exploratory design was not preregistered. No outcome-guided variable search, latent-factor item deletion, imputation or institutional weighting was performed. Covariate coefficients and sensitivity checks are not independent confirmatory replications. Source numerical values were not altered. Programme outputs were retained with the analytical script for reproducibility.