Journal of State Government and Mass Media

Volume: 4 Issue: 4
October-December, 2026



KMF PUBLISHERS

www.kmf-publishers.com/jsgmm/

Journal of State Government and Mass Media

Vol 4 Issue 4 (2026)

Research Article

OPEN

DOI: https://doi.org/10.64907/xkmf.v04i04.jsgmm.3

Deepfake Awareness and Self Reported Verification among AI Aware Respondents in Bangladesh

Md. Mirajul Islam[1*]; Mahmudul Hasan[1]; Nusrat Zahan Nishi[1]; Kazi Abdul Mannan[2]

Deepfake education is often expected to encourage verification, yet awareness, beliefs about intention and actual checking require different evidence. This study examines these distinctions in 385 adult, AI-aware survey records from Bangladesh, including a descriptive subset of 227 law students. The questionnaire separately asks about awareness of convincing AI-generated media, source checking, corroboration, the perceived influence of awareness on verification intention, and intended restraint in sharing unverified material. Responses were analysed using item distributions, joint endorsement, paired comparisons, exploratory rank correlations and threshold sensitivity summaries. Overall, 244 respondents (63.4%) endorsed deepfake awareness, 265 (68.8%) reported source checking, and 278 (72.2%) reported seeking corroboration. The statement that awareness increases verification intention received 327 endorsements (84.9%), exceeding source-checking endorsement by 16.1 percentage points (pointwise bootstrap 95% interval 10.1-21.8). However, awareness had small ordinal correlations with checking and corroboration (ρ = 0.084 and 0.077); neither was statistically distinguishable from zero after adjustment within the four-test family. Law-student summaries showed a similar ordering of endorsement. These findings concern differently worded self-reports and do not establish an intention-behaviour gap, intervention effect or inability to identify deepfakes. The article contributes a measurement-based distinction between endorsing an awareness-to-intention proposition and independently observing verification. It recommends evaluating media-literacy education through source tracing, corroboration and discernment tasks that include authentic as well as manipulated content. Undocumented recruitment and the mixed sample limit population inference.

Keywords: deepfakes; verification; media literacy; source credibility; behavioural intention; law students; Bangladesh

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

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

*Corresponding author: Md. Mirajul Islam, Email: rimondewan77@gmail.com
Copyright: © 2026 by the authors. Licensee KMF Publishers (www.kmf-publishers.com). This open-access article is distributed under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Convincing synthetic images, audio and video complicate an everyday question: what evidence should a person seek before trusting or sharing digital material? The answer is not simply to distrust everything produced with artificial intelligence (AI). Synthetic media can be used transparently for legitimate purposes, while authentic recordings can be presented with misleading captions or taken out of context. Effective verification requires attention to a claim, its source and corroborating evidence, rather than a binary assumption that synthetic means false and authentic means true.

Deepfakes can nevertheless create distinctive risks when they impersonate people or depict events that did not occur. Legal scholarship has examined implications for privacy, reputation and democratic accountability, including the possibility that genuine evidence becomes easier to dismiss as fabricated.[1] Experimental work has also found that deceptive synthetic political video can increase uncertainty and affect trust in news.[2] Those findings motivate verification research, but they do not establish what any respondent in the present Bangladesh sample has seen or experienced.

Education about the existence of deepfakes is therefore relevant but incomplete. Knowing that AI can create convincing media does not demonstrate the ability to identify a particular fabrication, locate the original source or recognise when evidence is insufficient. Köbis and colleagues' experiment illustrates the difference between familiarity, confidence and demonstrated detection performance.[3] Research comparing human and machine detection further shows that performance depends on the task and the material presented.[4] General awareness should consequently be distinguished from the actions and outcomes that education seeks to improve.

The study examines deepfake awareness and intention to verify online information in a Bangladesh sample with a law-student subset. The supplied Group 14 workbook contains 385 records, but only 227 are explicitly classified as law students; the remainder are law graduates, advocates or legal professionals, and other relevant respondents. The study therefore reports a mixed sample and a separate law-student subset. It does not assume that every respondent is currently enrolled at a university or that the students represent the national university population.

The measurement issue is equally important. The five focal questionnaire statements ask about awareness, reported source checking, reported corroboration, an attributed influence of awareness on intention, and conditional restraint in sharing. One item explicitly states the relationship of interest: "Deepfake awareness increases my intention to verify online information." Agreement expresses the respondent's view about that proposition; it does not independently demonstrate an effect of awareness. An analysis that combines all five statements into one score would make the proposed relationship difficult to interpret.

[1] Bobby Chesney and Danielle Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 California Law Review 1753.

[2] Cristian Vaccari and Andrew Chadwick, 'Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News' (2020) 6(1) Social Media + Society, doi:10.1177/2056305120903408.

[3] Nils C Köbis, Barbora Doležalová and Ivan Soraperra, 'Fooled Twice: People Cannot Detect Deepfakes but Think They Can' (2021) 24 iScience 103364.

[4] Matthew Groh and others, 'Deepfake Detection by Human Crowds, Machines, and Machine-Informed Crowds' (2022) 119 Proceedings of the National Academy of Sciences e2110013119; Matthew Groh and others, 'Human Detection of Political Speech Deepfakes across Transcripts, Audio, and Video' (2024) 15 Nature Communications 7629.

The central problem is thus how to examine a plausible educational proposition without confusing agreement with it for evidence that it operates. This study keeps the items separate and compares the distribution of responses to the awareness-to-intention statement with responses about source checking and corroboration. It also examines whether an independently worded awareness item is associated with those checking statements. These are complementary descriptive questions rather than tests of an educational intervention.

Existing research provides evidence on specific detection tasks and media-literacy interventions, but such evidence is not interchangeable with cross-sectional agreement statements. Accuracy prompts, for example, have been tested through experimental outcomes involving the quality of information people consider sharing.[5] Work on AI-generated visual misinformation has tested particular instructional materials against discernment outcomes and also examined changes in belief in authentic content.[6] These studies suggest that an educational evaluation should identify what changes and how it is measured.

The contribution here is a focused comparison of a self-attributed intention statement with separate reports of verification practices. The analysis does not assume that higher endorsement of intention than practice proves a psychological gap: the statements differ in wording, reference conditions and time frame. Its contribution is instead to show which distinctions are necessary before an awareness-to-verification claim can be evaluated. It combines paired response patterns, independent item associations and a law-student restriction without fitting an unsupported latent model.

The focus remains on deepfake awareness and verification rather than treating trust in AI-generated news, regulatory preferences or legal awareness as substitutes for checking. Those issues can inform wider research on misinformation, but each requires separate measures. The contribution is a clearly delimited analysis of how the five response objects relate within the recorded sample.

The study describes the five focal items, examines how awareness and reported checking coexist, compares selected endorsement proportions within respondents, and explores ordinal associations between awareness and the other statements. It also examines whether the descriptive ordering appears among recorded law students and under a stricter response threshold.

The questions are: How frequently are awareness, checking, corroboration, attributed intention and conditional sharing restraint endorsed? What joint response patterns occur? How large are selected within-person endorsement differences? What are the estimated awareness associations and their uncertainty? Do the descriptive patterns persist in the law-student subset? The answers concern the supplied records, not population prevalence or the effectiveness of a training programme.

2. Literature Review and Theoretical Foundation

The theoretical foundation links information credibility, media literacy and planned behaviour while preserving distinctions among familiarity, intention, self-report and performance. These distinctions guide the analysis and its limits.

[5] Gordon Pennycook and others, ‘Shifting Attention to Accuracy Can Reduce Misinformation Online’ (2021) 592 Nature 590; Gordon Pennycook and David G Rand, ‘Accuracy Prompts Are a Replicable and Generalizable Approach for Reducing the Spread of Misinformation’ (2022) 13 Nature Communications 2333.

[6] Sean Guo, Briony Swire-Thompson and Xiaoqing Hu, ‘Specific Media Literacy Tips Improve AI-Generated Visual Misinformation Discernment’ (2025) 10 Cognitive Research: Principles and Implications 38.

2.1 Deepfake awareness and verification outcomes

Awareness that AI can produce convincing images, audio and video is a general proposition. It does not establish whether a person has encountered a deepfake, how often they do so or whether they can detect one. A performance task requires identifiable stimuli and an evaluation criterion. A verification task may instead ask the participant to trace an upload, compare accounts or locate independent evidence. A person can fail to detect a fabrication visually while still reach a sound conclusion through corroboration.

The distinction is supported by research that uses different modalities and testing conditions. Groh and colleagues' 2024 study investigates detection across transcripts, audio and video, demonstrating the importance of specifying what information participants receive.[7] Its task-specific results cannot establish a universal detection ability or the performance of respondents in Bangladesh. The present questionnaire contains no stimulus task, so the article uses "awareness" and "reported checking" only in relation to the actual statements.

In the same way, an intention to avoid sharing unverifiable content is not an observed decision to withhold it. A person may not encounter the relevant situation, may lack time to check, or may interpret verification differently. These are possible explanations for future study, not variables measured here. The survey can describe how participants answered, but it cannot identify which process explains their response.

2.2 Information credibility and source evaluation

Credibility research examines how users judge information and the sources that provide it. Metzger and Flanagin discuss cognitive heuristics in online credibility assessment, where users may rely on cues rather than conduct exhaustive evaluation.[8] Familiarity, professional presentation or social endorsement may be convenient cues, but none guarantees that a specific claim is accurate. Credibility is therefore better evaluated through a combination of source assessment and evidence than inferred from appearance alone.

The source-checking item in this survey does not define how a source is checked. Looking at the name of an account is different from locating an original recording or investigating who is responsible for a website. Similarly, corroboration can involve apparently separate accounts that repeat the same unverified claim. A credible verification task should distinguish independent evidence from repetition. The questionnaire does not provide enough detail to score the quality of respondents' reported practices.

Wineburg and McGrew's work on lateral reading examines strategies used to evaluate digital information by consulting material beyond the original page.[9] It provides an operational example for teaching source evaluation, rather than evidence that participants in this study use that strategy. An educational programme could evaluate whether students identify the original source and compare independent evidence, while allowing "not

[7] Groh and others, 'Human Detection' (n 4).

[8] Miriam J Metzger and Andrew J Flanagin, 'Credibility and Trust of Information in Online Environments: The Use of Cognitive Heuristics' (2013) 59 Journal of Pragmatics 210.

[9] Sam Wineburg and Sarah McGrew, 'Lateral Reading and the Nature of Expertise: Reading Less and Learning More When Evaluating Digital Information' (2019) 121(11) Teachers College Record 1.

enough evidence" as a legitimate conclusion.

2.3 Media literacy and the limits of general warnings

Media literacy involves accessing, analysing, evaluating and acting on media in context. Hobbs's framework treats it as a set of practices rather than a narrow ability to identify one technical artefact.[10] AI literacy likewise extends beyond naming a technology to considering its capabilities, limitations and responsible use.[11] In this study, knowing that deepfakes exist is one measured self-assessment within that broader space.

Intervention evidence should be interpreted in relation to the materials, audience and outcome tested. Guo, Swire-Thompson and Hu compared specific and general media-literacy tips for AI-generated visual misinformation. Their study found differences in discernment and also reduced belief in genuine headlines relative to control conditions.[12] This illustrates why reducing belief in false content alone is not sufficient evidence of improved information judgement: indiscriminate scepticism can also reduce acceptance of accurate material.

Huang and Hu's two-wave experiment examined warning messages and technique tips in different presentation formats.[13] The study distinguishes detection, perceived literacy and self-efficacy rather than assuming that they are one outcome. The present questionnaire cannot replicate that design or establish which type of instruction would work locally. It can identify why future evaluations should keep confidence, intention and task performance separate.

2.4 Theory of planned behaviour

The theory of planned behaviour relates intentions to attitudes, subjective norms and perceived behavioural control, with behaviour depending partly on control over the action.[14] It offers a useful account of why recognising a risk may be insufficient to produce checking. Someone may favour accuracy while facing practical constraints or competing motivations. However, neither the existence of this theory nor agreement with a proposition about intention demonstrates that a specific pathway has operated in the present sample.

The instrument does not separately measure all planned-behaviour constructs, use a common prospective time window or observe subsequent checking. DVI4 attributes an increase in intention to deepfake awareness within a single sentence. DVI2 and DVI3 report practices in general terms, while DVI5 expresses a conditional willingness to refrain from sharing. Their contrast should not be labelled an observed intention-behaviour gap or used to estimate the size of such a gap.

A stronger test would measure awareness independently, specify a particular verification action and time frame, assess perceived control and social expectations, and observe a later opportunity to act. Researchers could then test whether intentions predict checking under those conditions. The current article uses planned-behaviour theory to clarify what is missing, rather than claim confirmation of an unmeasured model.

[10] Renee Hobbs, Digital and Media Literacy: A Plan of Action (Aspen Institute 2010).

[11] 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.

[12] Guo, Swire-Thompson and Hu (n 6).

[13] Guanxiong Huang and Bo Hu, '"A Warning Is Not Enough. Teach Me How to Spot Deepfakes.": Testing Media Literacy Interventions for Combating Deepfakes' (2025) Science Communication, advance online publication, doi:10.1177/10755470251382889.

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

2.5 Attention, detection and calibrated trust

Awareness and attention are distinct. Accuracy-prompt studies shift attention toward accuracy at a decision point and evaluate subsequent responses.[15] Those findings do not imply that respondents with higher general awareness necessarily check more frequently. A prompt is an intervention, whereas the present awareness item is a self-report obtained alongside other statements.

Detection is another distinct outcome. Köbis and colleagues found that awareness and incentives did not reliably improve detection under the conditions of their experiment.[16] Conversely, educational and inoculation studies show that some interventions can improve particular forms of misinformation resilience.[17] These findings need not conflict: interventions, content and outcome measures differ. A defensible local evaluation would specify whether it seeks better detection, better sourcing, more appropriate sharing or more calibrated confidence.

Calibrated trust means responding to the evidence rather than accepting or rejecting media indiscriminately. An authentic video can still communicate a false claim, while a synthetic clip may be clearly disclosed satire or illustration. Educators should therefore test both acceptance of accurate information and rejection of misleading content. The present questionnaire cannot estimate false-positive detection errors or the loss of trust in genuine reporting.

2.6 Media law and expression in Bangladesh

Deepfakes raise possible issues involving identity, privacy, reputation and deceptive uses, but the legal character of any item depends on its content, purpose, consent and applicable law. Chesney and Citron analyse a range of harms and responses while recognising legitimate uses and limitations of proposed remedies.[18] Their analysis is comparative scholarship, not a determination that a particular respondent's activity is unlawful in Bangladesh.

Article 39 of Bangladesh's Constitution guarantees freedom of thought and conscience and, subject to specified reasonable restrictions imposed by law, freedom of speech, expression and the press.[19] International guidance on expression emphasises the need to assess legality, necessity and proportionality when restricting expression.[20] These principles support careful policy design rather than treating uncertainty about authenticity as a sufficient basis to suppress content.

The survey does not record actual incidents, alleged offences, platforms' moderation decisions or participants' understanding of a specific legal provision. Its results cannot establish the prevalence of unlawful conduct or support for a particular legal restriction. Legal implications are therefore confined to the need for accurate information, proportionate safeguards and procedures for contesting mistakes. This article is not a comprehensive review of current Bangladeshi cyber or media legislation.

[15] Pennycook and others (n 5); Pennycook and Rand (n 5).

[16] Köbis, Doležalová and Soraperra (n 3).

[17] Jon Roozenbeek and others, 'Psychological Inoculation Improves Resilience against Misinformation on Social Media' (2022) 8 Science Advances eabo6254.

[18] Chesney and Citron (n 1).

[19] The Constitution of the People's Republic of Bangladesh, art 39.

[20] UN Human Rights Committee, 'General Comment No 34: Article 19, Freedoms of Opinion and Expression' (12 September 2011) UN Doc CCPR/C/GC/34, paras 22–36.

2.7 Analytical framework and questions

Figure 1 distinguishes the measured response types from outcomes not observed. DVI1 is a self-assessment of awareness; DVI2 and DVI3 report checking practices; DVI4 endorses an attributed influence on intention; and DVI5 states conditional restraint. None measures correct detection or observed verification. The figure is an analytical map, not an estimated causal model.

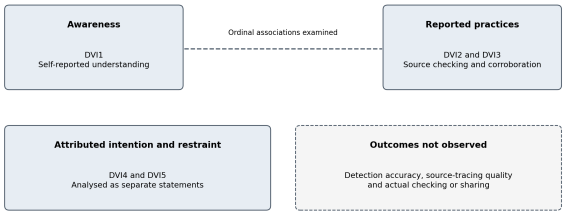

Figure 1. Analytical distinctions among measured responses and unmeasured outcomes

The dashed connection denotes an exploratory association, not a causal pathway. Attributed intention is analysed separately from awareness and practice reports. Detection accuracy and observed checking are not measured.

The primary paired comparisons are DVI4 versus DVI2, DVI4 versus DVI3, and DVI5 versus DVI2. Four ordinal associations compare DVI1 with DVI2-DVI5. They are exploratory questions selected for the current analysis, not preregistered directional hypotheses. The design permits a precise distinction: a large endorsement percentage for the attributed-intention statement can coexist with a small awareness-practice association without proving either a causal mechanism or its absence.

3. Methodology

The study uses a quantitative cross-sectional design. Reporting distinguishes recorded procedures from undocumented fieldwork information, consistent with transparency principles for observational studies.[21] Item-level analysis replaces latent-variable modelling because the focal statements have different meanings and do not independently operationalise all proposed theoretical constructs.

3.1 Design and recorded population

The analysis uses 385 records from the Group 14 workbook. All records are coded as consenting, at least 18 years old and aware of AI. Respondent status identifies 227 law students, 81 law graduates, 47 advocates or legal professionals, and 30 other relevant respondents. Current university enrolment is not established for every record; the law-student category provides the directly identifiable student subset.

The unit of analysis is an individual recorded response. No deepfake stimulus was presented in the supplied materials for this analysis, no task accuracy was recorded and no platform behaviour was observed. The study consequently estimates response distributions and associations within the supplied records. It does not evaluate an intervention, a detector or the prevalence of deepfake exposure.

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

3.2 Sampling and data provenance

The research plan specifies a target of 385 completed responses for a large or unknown population and prefers probability sampling. A conventional sample-size calculation can motivate that target under stated assumptions, but does not establish that random selection occurred.[22] The supplied materials do not document fieldwork dates, recruitment channels, a sampling frame, invitation counts, response rate or exclusions. No recruitment or administration procedure is inferred from the target or final row count.

The researcher identifies the numerical responses as real survey data. The Codebook nevertheless labels Respondent_ID as a synthetic respondent sequence and the optional contextual-text field as a synthetic comment. Those labels do not alone establish that numerical answers are synthetic; their creation and relationship to original questionnaires require documentation. The analysis excludes contextual text. Source records, consent documentation and recruitment information should be retained for verification and reproducibility.

3.3 Measures and content validity

All five focal items use the same ordered response scale: strongly disagree, disagree, neutral, agree and strongly agree. Table 1 specifies their wording and the interpretation supported by each. The common scale does not make the objects of measurement equivalent. In particular, DVI4 contains the relationship of interest in its wording and is not used as an independent measure of a causal effect.

Table 1. Item meanings and boundaries of inference

Item	Statement focus	Supported interpretation
DVI1	Awareness of convincing AI-generated media	Self-reported understanding; not demonstrated detection
DVI2	Checking a source before sharing suspicious or sensational content	Reported practice; no observed source-tracing task
DVI3	Seeking corroboration under authenticity uncertainty	Reported practice; independence and quality not tested
DVI4	Awareness increases intention to verify	Endorsement of an attributed influence; not an independently estimated effect
DVI5	Avoid sharing what cannot reasonably be verified	Conditional intention; no observed sharing choice

The focal block mixes different response objects. Exact wording is reproduced in Appendix A.

The article does not form a five-item “deepfake awareness” or “verification intention” index. Preliminary coefficient alpha for the whole DVI block is 0.341; even a higher value would not by itself establish construct validity or justify mixing self-assessment, report and attributed intention.[23] No adapted scale source, separate pilot dataset or validated measurement model is established by the supplied documentation. SEM loadings, AVE, HTMT and structural paths are not reported.

[22] William G Cochran, Sampling Techniques (3rd edn, John Wiley & Sons 1977).
[23] Klaas Sijtsma, ‘On the Use, the Misuse, and the Very Limited Usefulness of Cronbach’s Alpha’ (2009) 74 Psychometrika 107.

3.4 Screening and descriptive summaries

The workbook contains 385 unique identifiers and no missing values in DVI1-DVI5. Sixteen records lack the optional AI-tool response; that field is not used in the focal analyses. No exact duplicate records are present when the identifier is excluded, but this does not establish that the records correspond to independently recruited participants. No attention-check or time-to-completion variable is available for independent assessment.

The analysis retains all 385 focal records and applies no weights or imputation. Counts for every response category are reported. Endorsement means agree or strongly agree. Neutral responses remain visible in the distribution and are grouped with disagreement only as non-endorsement for joint and paired summaries. They are not classified as opposition. Pointwise Wilson intervals summarise endorsement uncertainty under conventional assumptions, without implying probability-sample national coverage.[24]

3.5 Joint responses and paired comparisons

Two joint summaries examine DVI1 with DVI2 and DVI4 with DVI2. Each divides the sample into four exhaustive response combinations: both endorsed, first only, second only and neither. They describe answer combinations rather than latent types of people. A first-only category includes neutral or disagreement responses to the second statement.

Three paired contrasts compare DVI4 with DVI2, DVI4 with DVI3, and DVI5 with DVI2. Exact McNemar tests use discordant pairs; Holm adjustment covers the three p-values as one family.[25] Pointwise percentile bootstrap intervals for endorsement differences use 10,000 whole-respondent resamples, preserving the pairing.[26] Different wording, reference conditions and response style can affect a contrast, so it is not interpreted as a causal effect, priority ranking or measured intention-behaviour gap.

3.6 Ordinal associations and sensitivity analyses

Four Spearman correlations compare the original five-category DVI1 responses with DVI2-DVI5. These retain ordinal information rather than correlate only endorsement indicators. Holm adjustment covers the four p-values as a separate family. Bootstrap intervals use 3,000 whole-respondent resamples and are pointwise, not multiplicity-adjusted. The analysis was not preregistered; estimates, intervals and adjusted p-values are reported together. Non-significance does not prove the absence of an association.[27]

The 227 law-student records are described separately, and the same three paired contrasts are repeated as an exploratory sensitivity analysis with their own three-test Holm family. This does not establish a difference between students and non-students. A stricter threshold, strongly agree only, is also tabulated descriptively for the full sample and student subset. It examines dependence on the response cut-off, not a separate validated construct. Sensitivity results are not an independent replication because the samples overlap.

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

[25] 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.

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

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

3.7 Ethical considerations

Ethical approval for this study was obtained from the Center for Academic & Professional Career Development and Research (CAPCDR), an external institutional body, under approval reference CAPCDR: SMUCT/LLB/SS-0020260908-23. External ethical review was sought because Shanto-Mariam University of Creative Technology does not have an institutional ethics committee responsible for reviewing this type of research.

4. Results

Results are organised around the distinctions in Table 1. They concern recorded responses and should not be interpreted as national prevalence, objective deepfake detection ability or observed verification.

4.1 Respondent profile

Table 2 describes recorded status, age and AI-use frequency. These categories delimit the sample and identify the directly recorded law-student subset.

Table 2. Recorded respondent profile

Measure	Category	Count (%)
Status	Law student	227 (59.0%)
	Law graduate	81 (21.0%)
	Advocate or legal professional	47 (12.2%)
	Other relevant respondent	30 (7.8%)
Age	21-23	140 (36.4%)
	24-26	88 (22.9%)
	18-20	69 (17.9%)
	31+	50 (13.0%)
	27-30	38 (9.9%)
AI-use frequency	Weekly	165 (42.9%)
	Daily or almost daily	105 (27.3%)
	Monthly	66 (17.1%)
	Less than monthly	38 (9.9%)
	Never	11 (2.9%)

N = 385. The law-student subset contains 227 records. Status does not establish representative sampling or current enrolment for other categories.

Law students form 59.0% of the sample (227/385); 81 records are law graduates, 47 advocates or legal professionals, and 30 other respondents. The largest age group is 21-23 (140; 36.4%). Weekly AI use is the most frequent category (165; 42.9%), while 11 respondents report never using AI. Awareness eligibility and use frequency are different variables, so awareness does not imply use.

4.2 Item distributions

Table 3 presents the five-category distributions and endorsement intervals. DVI1 concerns reported understanding, DVI2-DVI3 reported checking practices, DVI4 an attributed influence on intention, and DVI5 conditional restraint.

Table 3. Focal item distributions and endorsement

Item	Counts 1 / 2 / 3 / 4 / 5	Endorsed n (%)	95% CI (%)
DVI1	3 / 21 / 117 / 186 / 58	244 (63.4%)	58.5 to 68.0
DVI2	0 / 13 / 107 / 191 / 74	265 (68.8%)	64.0 to 73.3
DVI3	2 / 17 / 88 / 194 / 84	278 (72.2%)	67.5 to 76.4
DVI4	0 / 5 / 53 / 173 / 154	327 (84.9%)	81.0 to 88.2
DVI5	0 / 7 / 62 / 155 / 161	316 (82.1%)	77.9 to 85.6

N = 385. Categories 1-5 run from strongly disagree to strongly agree. Endorsement combines agree and strongly agree. Wilson intervals are pointwise and do not establish national coverage.

Awareness was endorsed by 244 respondents (63.4%); source checking by 265 (68.8%); and corroboration by 278 (72.2%). DVI4 received 327 endorsements (84.9%), and DVI5 received 316 (82.1%). These percentages refer to statements with different content. The results do not show that respondents who endorse DVI4 have actually changed their behaviour because of awareness.

Neutral responses account for substantial portions of the awareness and practice items:

117 for DVI1, 107 for DVI2 and 88 for DVI3. Their meaning cannot be inferred from the scale alone. Neither a neutral response nor non-endorsement is evidence that a respondent believes all online media or never checks a source.

4.3 Joint endorsement patterns

Table 4 shows joint response combinations for awareness and source checking, and for attributed intention and source checking. The denominators are 385 throughout.

Table 4. Joint endorsement of awareness, attributed intention and source checking

Item pair	Response combination	Count (%)
DVI1 / DVI2	Both	169 (43.9%)
	First only	75 (19.5%)
	Second only	96 (24.9%)
	Neither	45 (11.7%)
DVI4 / DVI2	Both	225 (58.4%)
	First only	102 (26.5%)
	Second only	40 (10.4%)
	Neither	18 (4.7%)

N = 385. First and second refer to the listed item pair. Neither and single-item categories include neutral responses to the non-endorsed item. Categories do not identify stable respondent types.

Awareness and source checking were both endorsed by 169 respondents (43.9%); 75 endorsed awareness without endorsing checking, while 96 endorsed checking without endorsing awareness. These responses do not identify whether anyone correctly recognised a deepfake. In the second comparison, 225 endorsed DVI4 and DVI2, 102 endorsed DVI4 only, 40 endorsed DVI2 only and 18 endorsed neither. A DVI4-only response is not treated as demonstrated failure to act on an intention because the items lack a shared prospective action and time frame.

4.4 Paired endorsement contrasts

Table 5 reports discordant counts, differences, pointwise intervals and adjusted tests. Each contrast uses the same individuals on both items.

Table 5. Paired endorsement contrasts

Contrast	First only / second only	Difference pp	95% CI pp	Holm p
DVI4 - DVI2	102 / 40	16.1	10.1 to 21.8	<0.001
DVI4 - DVI3	84 / 35	12.7	7.3 to 18.2	<0.001
DVI5 - DVI2	92 / 41	13.2	7.5 to 19.0	<0.001

N = 385; pp = percentage points. Exact McNemar tests; Holm adjustment across three contrasts. Intervals use 10,000 paired bootstrap resamples and are pointwise. Wording differs across items.

DVI4 endorsement exceeded DVI2 by 16.1 percentage points (95% interval 10.1-21.8) and DVI3 by 12.7 points (7.3-18.2). DVI5 exceeded DVI2 by 13.2 points (7.5-19.0). All three exact paired tests had Holm-adjusted p-values below 0.001. The tests establish differences in response distributions under the specified categorisation, not psychological mechanisms or intervention effects.

4.5 Awareness associations

Table 6 uses all five response categories to examine awareness associations. Adjusted p-values concern the four-test family; intervals are pointwise.

Table 6. Exploratory ordinal associations with awareness

Item pair	Spearman ρ	95% CI	p	Holm p
DVI1 / DVI2	0.084	-0.015 to 0.179	0.099	0.298
DVI1 / DVI3	0.077	-0.026 to 0.177	0.129	0.298
DVI1 / DVI4	0.097	-0.004 to 0.197	0.057	0.230
DVI1 / DVI5	0.067	-0.033 to 0.164	0.187	0.298

N = 385. Holm adjustment covers four tests. Percentile intervals use 3,000 whole-respondent bootstrap resamples and are pointwise, not multiplicity-adjusted.

The awareness-source-checking correlation was 0.084 (95% interval -0.015 to 0.179; Holm p = 0.298). Awareness-corroboration was 0.077 (-0.026 to 0.177; Holm p = 0.298). The other two correlations were also small: 0.097 with DVI4 and 0.067 with DVI5. None met the 0.05 threshold after adjustment. The estimates do not establish zero effects or show that awareness education would be ineffective.

4.6 Law-student and response-threshold sensitivity

Table 7 reports the law-student proportions and strongly-agree-only summaries. All denominators are stated to prevent the student subset from being mistaken for the full sample.

Table 7. Student restriction and stronger response threshold

Item	Students endorse n (%)	Student 95% CI (%)	All strongly agree n (%)	Students strongly agree n (%)
DVI 1	142 (62.6%)	56.1 to 68.6	58 (15.1%)	36 (15.9%)
DVI 2	154 (67.8%)	61.5 to 73.6	74 (19.2%)	41 (18.1%)
DVI 3	160 (70.5%)	64.2 to 76.0	84 (21.8%)	43 (18.9%)
DVI 4	190 (83.7%)	78.3 to 87.9	154 (40.0%)	91 (40.1%)
DVI 5	193 (85.0%)	79.8 to 89.1	161 (41.8%)	93 (41.0%)

All sample N = 385; law students n = 227. Endorse means agree or strongly agree. Student intervals are pointwise Wilson intervals. Strongly-agree summaries are descriptive; subsets overlap.

Among 227 law students, awareness was endorsed by 142 (62.6%), source checking by 154 (67.8%), corroboration by 160 (70.5%), DVI4 by 190 (83.7%) and DVI5 by 193 (85.0%). The broad descriptive ordering therefore resembles the full sample. The three student-only paired differences were 15.9, 13.2 and 17.2 percentage points respectively; their separately adjusted p-values were below 0.001. These overlapping-sample results do not constitute an independent replication or test between populations.

With strongly agree alone, 58 full-sample respondents endorsed DVI1 at that level, compared with 74 for DVI2, 84 for DVI3, 154 for DVI4 and 161 for DVI5. The descriptive separation between the attributed-intention item and source checking thus remains under the stricter threshold. The change in cut-off reduces the counts and changes the estimand; no new inferential claim is made from this sensitivity summary.

5. Discussion

The principal result is a distinction between endorsing a proposition about awareness and separately reporting verification practices. DVI4 attracted high agreement, while awareness associations with independently worded checking statements were small and uncertain. These findings answer different questions and are not contradictory.

5.1 Endorsing a relationship is not demonstrating it

Agreement that deepfake awareness increases verification intention expresses respondents' understanding of their own motivation. It may be informative for designing educational messages, but it is not a direct measure of what changes when awareness increases. The survey neither manipulates awareness nor follows respondents over time. Treating the DVI4 percentage as evidence that an awareness intervention works would confuse the content of a belief with a measured effect.

The paired contrasts show that attributed intention and conditional restraint received more endorsement than reports of checking or corroboration. This does not prove that respondents fail to implement an intention. They may interpret the questions differently, imagine different situations, or express stronger agreement with a conditional principle than with a general report of practice. The dataset does not identify which explanation applies. Planned-behaviour theory makes a later intention-action test plausible, but the present measures do not supply it.[28]

The joint responses are useful because they preserve coexistence without imposing a single score. For example, the 102 DVI4-only records show agreement with the attribution statement alongside non-endorsement of the checking item. Those records identify a group for follow-up questions about interpretation and opportunity to check. They do not justify labelling participants inconsistent, careless or deceptive.

5.2 Small associations and measurement limits

The awareness-checking and awareness-corroboration estimates were small, with intervals spanning negative and positive values. The analysis provides limited evidence about their association in the supplied sample. It does not establish that increasing awareness has no causal effect. Exposure to information, task difficulty, access to verification resources, response style and unmeasured motivations could influence the pattern; they were not assessed as mechanisms here.

A single awareness statement may also have limited ability to distinguish levels of knowledge. Some participants may understand the existence of synthetic media without recognising the term "deepfake"; others may strongly agree while lacking detection skills. The experiment by Köbis and colleagues demonstrates why awareness and measured ability need not coincide, but it should not be used to diagnose overconfidence in respondents who completed no accuracy task.[29]

The low alpha of the combined block is consistent with caution about aggregation, yet alpha is not the central reason for separate analysis. The substantive reason is that the items ask about different objects. A highly consistent set of responses would still not transform attributed intention into behaviour or establish the missing planned-behaviour constructs. Measurement should precede model selection rather than be stretched to accommodate a preferred SEM design.[30]

[28] Ajzen (n 14).
[29] Köbis, Doležalová and Soraperra (n 3).
[30] Sijtsma (n 23).

5.3 Verification beyond visual detection

The source-checking and corroboration statements suggest activities that can extend beyond inspecting pixels or listening for artefacts. A user can search for the earliest available source, compare the claim with independent reporting, check whether the material is presented in context and recognise when authenticity remains unresolved. Lateral-reading research offers examples of such evaluation outside the original page.[31]

This distinction matters because detection tools and visual cues can fail or become less useful as media-generation techniques change. Research comparing humans and machines illustrates that accuracy is contingent on the material and assistance provided.[32] The current survey does not establish how participants check sources, whether apparent corroboration is independent, or whether their decisions are correct. An educational programme should therefore assess the quality of verification, not only how often participants say they check.

A useful assessment would include authentic content alongside manipulated and misleadingly captioned content. It should record false accusations of manipulation as well as missed fabrications. Otherwise, an intervention that produces widespread distrust could appear effective merely because participants reject more content. That concern is supported by work on media-literacy tips showing changes in belief in real as well as false material.[33]

5.4 Implications for law students and educational design

The law-student subset has similar descriptive ordering to the full sample, but the workbook does not identify institutions, courses completed or prior deepfake training. These data cannot establish that legal education produces a particular awareness or practice level. Nor should the subset be taken as representative of university students in unrelated disciplines or of the public.

For students studying law, verification may include distinguishing the authenticity of a recording from the truth of a claim and from its legal relevance. These are separate questions. An authentic recording can be incomplete, while a clearly disclosed reconstruction may have a legitimate explanatory purpose. Training should ask what is being asserted and what evidence supports it, rather than classify all synthetic material as unlawful or all authentic media as reliable.

Experimental literature provides options for evaluation, not a ready-made local solution. Accuracy prompts, inoculation and targeted instruction have been tested through different tasks and populations.[34] A Bangladesh programme could compare methods using locally understandable materials, predefine outcomes and measure transfer to unfamiliar examples. This article does not select a superior intervention or claim that one will work without adaptation.

[31] Wineburg and McGrew (n 9).
[32] Groh and others, ‘Deepfake Detection’ (n 4).
[33] Guo, Swire-Thompson and Hu (n 6).
[34] Pennycook and Rand (n 5); Roozenbeek and others (n 17); Huang and Hu (n 13).

5.5 Media law and proportionate responses

Awareness of potential deception can justify educational and technical precautions, but uncertainty alone should not determine removal, punishment or the legal status of content. Article 39 and international expression principles require attention to rights and legally justified limits.[35] The study's checking statements do not indicate support for censorship, a particular statute or automated enforcement.

The appropriate response depends on the harm and evidence. Impersonation, reputational injury, fraudulent solicitation and disclosed creative use raise different questions; the survey does not identify any such incident. Comparative legal scholarship discusses the risks of both manipulated evidence and dismissal of genuine evidence.[36] That dual concern supports procedures for verification and correction that permit uncertainty and challenge, rather than equating suspicion with proof.

The policy implication is limited but concrete: institutions can improve access to evidence and verification support while preserving opportunities to contest errors. Decisions about specific legal restrictions require legal analysis and evidence beyond this questionnaire. No finding here establishes a new legal duty for individual users to verify every item they encounter.

5.6 Integrated interpretation

The three theoretical perspectives point toward a measured sequence of questions: what does the person know, how do they evaluate a source, what do they intend to do, what can they practically do, and what happens when they encounter a claim? The current survey covers self-reported portions of these questions and one attributed relationship. It does not validate a complete sequence or estimate mediation.

The distinctive finding is therefore evidential rather than causal. High endorsement of an awareness-to-intention statement coexists with small independent awareness-practice associations and lower endorsement of checking statements. This combination identifies an important measurement boundary and a useful direction for follow-up research. It does not establish a behavioural deficit, an education effect or the irrelevance of awareness.

6. Theoretical and Practical Implications

The following implications separate conclusions supported by the records from proposals for educational and policy evaluation. They do not describe interventions conducted in the present study.

6.1 Theoretical implications

First, research should distinguish endorsement of a causal-sounding proposition from an independently estimated relationship. If the outcome item already states that awareness increases intention, its association with an awareness item cannot by itself validate the proposed mechanism. Researchers should measure awareness and a specific prospective intention separately and observe later behaviour where possible.

[35] Constitution of Bangladesh (n 19); Human Rights Committee (n 20).
[36] Chesney and Citron (n 1).

Second, verification should be defined operationally. Source identification, lateral reading, independent corroboration, abstention from sharing and correct detection can be related while remaining different outcomes. Credibility and media-literacy perspectives help identify those practices, but a single composite should not erase their differences.[37]

Third, planned-behaviour theory should be tested using measures that match the action, context and time frame.[38] A follow-up model could include norms and perceived control only if they are actually measured. The present work contributes a clearer measurement specification and descriptive evidence, rather than confirmation or rejection of the complete theory.

6.2 Practical implications for universities and platforms

Universities could use short tasks in which students identify an original source, seek genuinely independent corroboration and explain what remains unknown. Assessment should reward calibrated conclusions rather than confidence alone. Task sets should contain real as well as manipulated material and should not rely on a fixed list of visual glitches as a timeless detection strategy.

At a practical decision point, platforms and educators could evaluate prompts that encourage checking a source or pausing before sharing. The accuracy-prompt literature provides experimental precedents, but local implementation needs testing for efficacy, accessibility and unintended distrust.[39] A high DVI4 endorsement percentage is not sufficient evidence to deploy one intervention or predict its effect.

Support should be accessible to people with different language needs, devices and levels of digital experience. The dataset does not measure these barriers, so the recommendation is to investigate them before designing a programme. Staff should also be able to explain how to handle uncertainty, correct a mistaken share and seek help without exposing unnecessary personal information.

6.3 Policy and evaluation implications

Institutions can distinguish user education from enforcement. Teaching source evaluation need not require collecting private browsing histories, and a research task can use benign examples rather than exposing participants to non-consensual or highly distressing synthetic content. A later study should provide consent, debriefing and privacy safeguards appropriate to its stimuli and logging methods.

Evaluation should report both missed misinformation and rejection of accurate information. It should also document the task, comparison condition and follow-up period. Legal responses must be assessed independently against applicable rights and proportionality requirements.[40] The present sample cannot establish public support for a regulatory programme or the population burden of deepfake harm.

[37] Metzger and Flanagin (n 8); Hobbs (n 10).
[38] Ajzen (n 14).
[39] Pennycook and others (n 5); Pennycook and Rand (n 5).
[40] Human Rights Committee (n 20).

7. Conclusion, Limitations and Future Research

In 385 AI-aware records from Bangladesh, 84.9% endorsed the statement that deepfake awareness increases verification intention, compared with 68.8% reporting source checking and 72.2% reporting corroboration. Paired endorsement differences were positive, while independent awareness-practice correlations were small and statistically uncertain after adjustment. The law-student subset showed a similar descriptive ordering. These findings demonstrate distinctions among questionnaire responses, not an intervention effect or an observed intention-behaviour gap.

The contribution is to keep awareness, attributed intention, checking reports and performance conceptually separate. It provides a focused empirical account of how they coexist and specifies the evidence required for stronger claims. A single self-reported attribution cannot substitute for a longitudinal or experimental test of whether awareness produces better verification.

Limitations include undocumented recruitment and dates, a mixed sample dominated by law students, unverified current university enrolment outside the student category, broad single-item measures, self-report and no detection or behavioural task. The numerical responses are identified by the researcher as real data; the Codebook's synthetic identifier and optional-text labels require documentation against original records. Sensitivity analyses reuse overlapping records and do not constitute independent replication. Sampling uncertainty does not resolve selection bias.

Future research should document recruitment and consent, recruit a defined student population, validate separate measures and preregister a limited analysis plan. A study could measure awareness before a realistic verification opportunity, record source-tracing choices and score correct acceptance, rejection and calibrated uncertainty. Training comparisons should test transfer to unfamiliar examples and follow-up retention while assessing potential distrust of authentic information. Interviews could clarify what participants mean by "checking a source" and what prevents them from doing so.

Declarations

Ethical approval and oversight

Ethical approval for this study was obtained from the Center for Academic & Professional Career Development and Research (CAPCDR), an external institutional body, under approval reference CAPCDR: SMUCT/LLB/SS-0020260908-23. 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 that informed consent was obtained before questionnaire completion. The questionnaire states that participation is voluntary, no name is required and participants may stop at any time. All analysed records contain affirmative consent and adult-eligibility responses.

Data availability

The numerical data supporting this analysis are held by the research team. Requests for de-identified data and analytical specifications may be directed to Prof. Dr Kazi Abdul Mannan. Any sharing is subject to participant confidentiality and applicable institutional requirements. Contextual comments were excluded.

Funding and competing interests

The authors report that the study received no external funding and declare no competing interests.

Author contributions

Md. Mirajul Islam: Conceptualisation; preparation of survey 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, manuscript drafting, language editing, organisation and development of analytical code. Numerical results were calculated programmatically from the supplied survey workbook. The authors are responsible for checking and approving the study's methods, findings, sources and interpretation in accordance with the target journal's policy.

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

Chesney B and Citron D, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019) 107 California Law Review 1753

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

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

Groh M and others, 'Deepfake Detection by Human Crowds, Machines, and Machine-Informed Crowds' (2022) 119 Proceedings of the National Academy of Sciences e2110013119

Groh M and others, 'Human Detection of Political Speech Deepfakes across Transcripts, Audio, and Video' (2024) 15 Nature Communications 7629

Guo S, Swire-Thompson B and Hu X, 'Specific Media Literacy Tips Improve AI-Generated Visual Misinformation Discernment' (2025) 10 Cognitive Research: Principles and Implications 38

Hobbs R, *Digital and Media Literacy: A Plan of Action* (Aspen Institute 2010)

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

Huang G and Hu B, '"A Warning Is Not Enough. Teach Me How to Spot Deepfakes.": Testing Media Literacy Interventions for Combating Deepfakes' (2025) Science Communication, advance online publication, doi:10.1177/10755470251382889

Köbis NC, Doležalová B and Soraperra I, 'Fooled Twice: People Cannot Detect Deepfakes but Think They Can' (2021) 24 iScience 103364

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

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

Metzger MJ and Flanagin AJ, 'Credibility and Trust of Information in Online Environments: The Use of Cognitive Heuristics' (2013) 59 Journal of Pragmatics 210

Pennycook G and others, 'Shifting Attention to Accuracy Can Reduce Misinformation Online' (2021) 592 Nature 590

Pennycook G and Rand DG, 'Accuracy Prompts Are a Replicable and Generalizable Approach for Reducing the Spread of Misinformation' (2022) 13 Nature Communications 2333

Roozenbeek J and others, 'Psychological Inoculation Improves Resilience against

Misinformation on Social Media' (2022) 8 Science Advances eabo6254

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

UN Human Rights Committee, 'General Comment No 34: Article 19, Freedoms of Opinion and Expression' (12 September 2011) UN Doc CCPR/C/GC/34, paras 22-36

Vaccari C and Chadwick A, 'Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News' (2020) 6(1) Social Media + Society, doi:10.1177/2056305120903408

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

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

Wineburg S and McGrew S, 'Lateral Reading and the Nature of Expertise: Reading Less and Learning More When Evaluating Digital Information' (2019) 121(11) Teachers College Record 1

Appendix A. Questionnaire wording

Table A1 reproduces the focal statements so that the meaning of each result can be assessed against its wording.

Table A1. Exact wording of the focal statements

Code	Questionnaire statement
DVI1	I understand that AI can create convincing fake images, audio and video.
DVI2	I check the source before sharing suspicious or sensational digital content.
DVI3	I look for corroborating evidence when media authenticity is uncertain.
DVI4	Deepfake awareness increases my intention to verify online information.
DVI5	I would avoid sharing content that I cannot reasonably verify.

Source: supplied Group 14 questionnaire and workbook Codebook. Each item uses a five-point agreement scale. The statements do not form a validated single scale.

The analysis retains the original responses and reports the items separately. DVI4 contains an attributed relationship within its wording.

Table of legislation

The Constitution of the People's Republic of Bangladesh, art 39