

Article

Digital Media Literacy Among Peruvian University Students: News Credibility in the Age of Manipulated Political Media

Robinson Bernardino Almanza Cabe ¹, Angela Aurora Mamani Calizaya ¹, Luis Dante Zubia Cortéz ¹, Jorge Erick Fernandez Corrales ¹, Alberto Limache Flores ¹ , Antonio Víctor Morales Gonzales ² , Adolfo Erick Donayre Sarolli ³, José Luis Chavez Cuarite ² and Miluska Odely Rodriguez Saavedra ^{4,*} 

¹ Faculty of Economics, Universidad Nacional de Moquegua, Moquegua 18001, Peru; ralmanzac@unam.edu.pe (R.B.A.C.); amamanic@unam.edu.pe (A.A.M.C.); lzubiac@unam.edu.pe (L.D.Z.C.); 20260201006@unam.edu.pe (J.E.F.C.); alimachef@unam.edu.pe (A.L.F.)

² Faculty of Economics, Universidad Nacional de San Agustín de Arequipa, Arequipa 04000, Peru; amoralesg@unsa.edu.pe (A.V.M.G.); jchavezcu@unsa.edu.pe (J.L.C.C.)

³ Graduate School, Faculty of Business Administration, Universidad Católica de Santa María, Arequipa 04013, Peru; adonayre@ucsm.edu.pe

⁴ Graduate School, Faculty of Law, Universidad Tecnológica del Perú, Arequipa 04001, Peru

* Correspondence: c28087@utp.edu.pe

Abstract

The rise in political deepfakes threatens news credibility and media literacy in polarized higher education contexts; yet, Latin American students remain understudied. This study examined the digital media literacy of Peruvian university students ($n = 5854$) in assessing the credibility of political news and identifying AI-generated content. Using a quantitative, non-experimental, cross-sectional design, PROCESS Models 4 and 5 were applied to analyze the relationship between AI literacy and news credibility assessment ability. Awareness of political deepfakes and political news verification strategies were examined as parallel mediators, while the type of university was evaluated as a moderator. Previous exposure to AI, digital media consumption, and sociodemographic characteristics were included as control variables. AI literacy was positively associated with news credibility assessment ability, with awareness of political deepfakes and verification strategies contributing to this relationship. The analysis also examined institutional differences between public and private universities. These findings support the implementation of systematic AI literacy programs across disciplines and institutions to strengthen news credibility and democratic resilience in politically polarized contexts.

Keywords: Digital media literacy; artificial intelligence; media literacy; AI literacy; misinformation; higher education; news credibility; manipulated political media



Academic Editor: Andreu Casero-Ripollés

Received: 6 April 2026

Revised: 3 July 2026

Accepted: 13 July 2026

Published: 17 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The proliferation of generative artificial intelligence technologies has radically transformed the global information ecosystem, giving rise to an era where the distinction between authentic and synthetic content has become increasingly complex (Sohaib & Tehseen, 2023; T. Wang et al., 2023). In this scenario, deepfakes (hyper-realistic videos, images, and audio generated using deep learning algorithms) represent one of the most sophisticated and potentially dangerous manifestations of this technological revolution (H. Wang et al., 2023). Their application in the political sphere is particularly worrying, given that this synthetic content can be used to manipulate public perceptions, exacerbate emotional polarization, and erode trust in democratic institutions (Pedersen et al., 2025; Tassone et al., 2024).

In the face of this threat, digital media literacy emerges as a necessary 21st-century skill, particularly for university students, who are not only intensive consumers of digital media, but also future professionals who will act as multipliers of information in their respective fields (Bhuvaneswari et al., 2025; Sharma et al., 2025). To address this challenge, the present study employs a quantitative, non-experimental, cross-sectional design using mediation and moderation models to examine how AI literacy shapes deepfake identification among Peruvian university students.

The Peruvian case has unique characteristics that make it a particularly vulnerable scenario and, paradoxically, one that has been little studied in relation to disinformation through deepfakes. In recent years, the country has experienced a profound political crisis characterized by multiple changes in government, massive social mobilizations, and marked citizen polarization, creating fertile ground for information manipulation (Garriga et al., 2024; Sánchez-Esparza et al., 2024; Rodríguez-Saavedra et al., 2026c). In this scenario, institutional communication (from presidential conferences to legislative statements) coexists with a fragmented media ecosystem where traditional media, digital platforms, and social networks proliferate without reliable verification mechanisms (Okolie, 2023; Simó-Soler, 2023), amid a broader process of digital transformation of Peruvian public institutions that has not been matched by equivalent advances in citizens' capacity to evaluate the authenticity of the digital content those institutions produce (Vilca Corrales et al., 2026). This fragmentation directly undermines news credibility: when citizens cannot distinguish authentic institutional statements from synthetic audiovisual content, trust in journalism and democratic information sources erodes progressively, a pattern consistent with broader institutional trust erosion documented in the Peruvian context (Vaccari & Chadwick, 2020; Morales Gonzales et al., 2026). Furthermore, the high penetration of mobile devices and social media among Peru's young population, combined with limited local fact-checking initiatives, creates a situation where potentially manipulated content can go viral before being verified (Duc Truong, 2024; Rodríguez-Saavedra et al., 2026b). In addition, as a developing country, Peru faces specific challenges in terms of digital literacy and unequal access to quality education in media skills (Battista, 2024). While Europe and North America have implemented systematic media literacy programs to combat deepfakes (Bhandari & Bhandari, 2025; Sanchez-Acedo et al., 2024), Latin America in general and Peru in particular lack curricular frameworks that prepare university students to identify and critically evaluate AI-generated content, even as generative AI tools are already being adopted across Peruvian public and administrative sectors (Rodríguez Saavedra et al., 2026). The threat is not merely hypothetical in the Peruvian context. Political deepfakes targeting public figures have circulated across Peruvian social media platforms during recent electoral cycles, manipulating the perceived statements of candidates and government officials in ways that directly distort public judgment about their credibility (Garriga et al., 2024). This pattern mirrors evidence from other electoral contexts, where campaign-related deepfakes have been shown to intensify affective polarization by shaping how voters evaluate opposing candidates (Neyazi et al., 2025). Unlike general misinformation, politically motivated deepfakes exploit the perceived authenticity of audiovisual evidence, making them particularly damaging to news credibility in societies where institutional trust is already fragile (Seker et al., 2025; Vaccari & Chadwick, 2020). For university students who consume political information primarily through social media, the inability to identify synthetic audiovisual content translates directly into vulnerability to news manipulation at a critical period of civic formation.

Despite the growing academic output on deepfakes and digital media literacy in Europe, Asia, and North America (Momeni, 2025; Vecchietti et al., 2025), there is a notable lack of research addressing this issue from the perspective of university students in Latin American countries, although emerging evidence on AI literacy and deepfake detection competence has begun to appear in the Peruvian higher education context (Rodriguez-Saavedra et al., 2026a). This absence is critical given that the sociocultural, political, and technological characteristics of the region (including varying levels of digital literacy, unequal access to educational resources, and volatile political contexts) shape specific experiences and vulnerabilities that cannot be directly extrapolated from studies conducted in the Global North (Hynek et al., 2025; Labuz, 2025; Rodriguez-Saavedra et al., 2026a). Therefore, contextualized research is needed to examine how Peruvian university students understand, identify, and respond to AI-generated synthetic content, as well as what critical skills they possess or lack to exercise informed and responsible digital citizenship in an increasingly complex and potentially manipulated media ecosystem (Bender, 2024; Celik, 2023).

Based on the above, the question guiding this study is: How does AI literacy relate to political news credibility assessment among Peruvian university students, and does this relationship differ between public and private universities? To answer this question, the following specific objectives are proposed: (1) to evaluate students' ability to distinguish authentic political news content from AI-generated deepfakes; (2) to determine the verification strategies applied to suspicious political content; (3) to identify the main vulnerabilities in political news assessment; and (4) to examine whether type of university moderates the relationship between AI literacy and news credibility assessment ability. These objectives will generate empirical evidence on the current state of Peruvian university students' digital skills in the face of AI manipulation.

In this sense, this research represents the first systematic exploration of digital media literacy skills for identifying deepfakes among Peruvian university students, thus filling a critical gap in Latin American academic literature on disinformation and media education (Musyaffi et al., 2024). Its findings will contribute to the development of contextualized curricular frameworks that enable Peruvian universities to train professionals capable of exercising critical thinking in complex digital environments, thus contributing to the construction of a more resilient citizenry in the face of information manipulation (Muawanah et al., 2024; Ofem et al., 2024). The study also offers relevant empirical evidence for educational policy makers, allowing them to base decisions on investment in digital literacy programs and curriculum updates in higher education, consistent with evidence that structured integration of AI tools into teaching practice strengthens digital literacy outcomes among university students (Kazanidis & Pellas, 2024; Yang, 2024; Zuo et al., 2025). By focusing on a developing country with specific political and media characteristics, this research provides valuable insights for other countries in the region facing similar challenges at the intersection of technology, journalism, and democracy (Saklaci & Gardikiotis, 2024). In doing so, it contributes directly to the emerging debate on how AI literacy can serve as a safeguard for news credibility and democratic resilience in the age of political deepfakes.

Literature Review

The intersection between generative artificial intelligence, deepfakes, and digital media literacy constitutes an emerging field of research with a marked concentration in the Global North and a notable scarcity of studies in Latin America.

Deepfakes are defined in this study as audiovisual content, namely videos, images, and audio recordings, generated or manipulated through deep learning algorithms to produce synthetic representations that are indistinguishable from authentic material (Mirsky & Lee, 2021). This definition deliberately excludes AI-generated text, which, although capable of spreading misinformation, lacks the indexical properties that make audiovisual content uniquely persuasive. The concept of indexicality, derived from semiotics, refers to the culturally embedded assumption that photographs and videos bear a direct causal relationship with the reality they depict, functioning as evidence rather than mere representation (Barthes, 1980/1981). It is precisely this indexical trust that deepfakes exploit: when a video appears to show a public figure speaking, audiences apply the same credibility they would grant to direct perceptual experience, making audiovisual synthetic content qualitatively more dangerous than written misinformation. A related conceptual question deserves acknowledgment: whether fully AI-generated audiovisuals that do not manipulate the likeness of a real person also qualify as deepfakes remains debated in the literature (Gronseth et al., 2025; Wu & Zhang, 2025). For the purposes of this study, both manipulated and fully synthetic audiovisuals are treated as deepfakes when their intent is to deceive about the authenticity of the content. The societal consequences of this technology are particularly acute in politically polarized contexts, where synthetic content can amplify social divisions, delegitimize political actors, and erode trust in journalism as an institution (Obateru et al., 2025; Seker et al., 2025). Existing regulatory frameworks have proven insufficient to address the pace of this technological evolution, creating conceptual and legal gaps that leave citizens, including university students, largely unprotected (Gronseth et al., 2025; Wu & Zhang, 2025).

Digital media literacy has evolved beyond a technical skill set into a broader framework of critical and ethical competencies needed to interpret AI-mediated information environments (Khan et al., 2025). A central tension in the literature concerns whether awareness of deepfakes alone is sufficient to protect audiences, or whether structured literacy training is required to translate awareness into reliable detection behavior. Evidence consistently supports the latter: passive warnings about deepfakes produce negligible improvements in identification accuracy, whereas programs that combine explicit instruction on technical markers with guided practice and metacognitive reflection generate meaningful gains (Zhao et al., 2025). This distinction matters because young people, despite expressing widespread skepticism toward online content, tend to rely on intuitive heuristics rather than systematic verification strategies when confronted with suspicious material (Bircan et al., 2025). The implication is that skepticism without structured competency does not reliably protect against deception. Restoring trust in news content therefore requires not only that audiences possess critical competencies to evaluate authenticity, but also that journalists and media institutions maintain transparent professional standards that give those competencies something reliable to work with (Lin et al., 2025). Adapting these frameworks to Latin American linguistic and cultural realities represents an additional challenge that the existing literature, predominantly produced in English-speaking contexts, has only begun to address (Özüdoğru & Durak, 2025).

Research on the digital competencies of university students reveals a persistent gap between technical exposure and critical capacity that is central to understanding why AI literacy training is necessary. Technical knowledge about AI systems does not automatically translate into the ability to critically evaluate their outputs, evidencing a disconnect between knowing how AI works and being able to detect when it is being used to deceive (B. Wang et al., 2025). This disconnect is compounded by psychological barriers: students with lower baseline digital skills experience technological anxiety that actively hinders the development of critical competencies, suggesting that digital literacy itself mediates the

relationship between technological exposure and critical development (Lao et al., 2025). AI literacy, understood as a distinct and multidimensional construct encompassing technical understanding, critical thinking about algorithmic biases, and evaluation of ethical implications, does not emerge spontaneously from general digital use but requires explicit and structured instruction (Huang & Hu, 2025; Sohaib & Tehseen, 2023). Disciplinary background further shapes how students engage with these competencies: engineering and computer science students tend to demonstrate stronger technical understanding, whereas students in education and humanities show greater critical capacity to evaluate the social and pedagogical implications of synthetic content, reinforcing the case for interdisciplinary approaches to AI literacy programs (Kaewsawad & Duda, 2025). Evidence from developing country contexts, including Indonesia, suggests that increasing risk awareness of AI-related threats can improve students' openness to engaging with literacy programs, a finding relevant to Latin American contexts that share comparable levels of institutional development and unequal digital access (Kontos et al., 2025).

The available evidence on educational interventions points toward a consistent model of how AI literacy develops and why its direction of influence on deepfake identification is theoretically grounded. Effective programs share four characteristics: cross-curricular integration, a combination of technical instruction with ethical reflection, opportunities for direct experimentation, and continuous updating as detection challenges evolve (Simó-Soler, 2023; Sharma et al., 2025). Critically, the causal direction assumed in this study, whereby AI literacy precedes and shapes deepfake awareness rather than the reverse, is supported by the logic of competency development: awareness of deepfakes as a risk is an outcome that requires prior conceptual grounding in how AI systems generate synthetic content, not a condition that spontaneously motivates individuals to seek technical understanding (Bhuvaneswari et al., 2025; Okolie, 2023). Put differently, individuals first need to understand what AI can do before they can meaningfully evaluate whether a specific piece of content might be synthetic. Active pedagogies that involve producing AI-mediated content, rather than passively consuming it, have demonstrated measurable improvements in critical thinking and self-efficacy, suggesting that learning through creation builds the conceptual schemas necessary to detect manipulation (Tassone et al., 2024). However, the mere incorporation of AI tools into curricula without intentional pedagogical design that promotes critical reflection fails to produce these gains, reinforcing that the quality of the instructional approach, not the technology itself, determines outcomes (Pedersen et al., 2025).

Implementing AI literacy programs in developing country contexts introduces structural barriers that shape what interventions can achieve. Infrastructure gaps, unequal access to digital devices, and heterogeneous teacher training mean universities in countries like Peru cannot simply import frameworks designed for the Global North (Duc Truong, 2024), while deficits from earlier educational stages force institutions to address foundational digital competencies before higher-order critical skills (Garriga et al., 2024). Low technological self-efficacy compounds these challenges, as it actively suppresses engagement with literacy interventions (Sánchez-Esparza et al., 2024), and AI integration can generate professional anxiety among educators that undermines program quality without adequate institutional support (Battista, 2024). Technically, specific audiovisual markers of synthetic content, such as lighting inconsistencies and artifacts in facial transitions, can be incorporated into literacy curricula (Bhandari & Bhandari, 2025), though human expert judgment remains necessary alongside automated systems, since pedagogical integration through guided practice determines whether students can apply this knowledge under real conditions of uncertainty (Sanchez-Acedo et al., 2024; Vecchiotti et al., 2025).

The review reveals three significant gaps. First, there is a scarcity of studies on digital media literacy and deepfakes in Latin American contexts. Second, research is concentrated in the Global North, which has different infrastructures. Third, few studies examine university students’ abilities to identify deepfakes in polarized political contexts. This research seeks to fill these gaps through an empirical study of Peruvian university students.

A fourth gap concerns the relationship between deepfake identification and news credibility: exposure to political deepfakes has been shown to erode trust in news media even among individuals who do not fully accept the fabricated content as true, a phenomenon particularly consequential in politically polarized contexts like Peru (Vaccari & Chadwick, 2020; Hancock & Bailenson, 2021). This study addresses all four gaps through an empirical examination of Peruvian university students.

2. Methodology

This research adopted a quantitative approach with a non-experimental cross-sectional design and explanatory scope. Mediation and moderation models were used to analyze the relationships between AI literacy and the ability to identify deepfakes in Peruvian university students, considering the effect of mediating and moderating variables (Labuz, 2025; Momeni, 2025). The two mediators (awareness of political deepfakes and verification strategies) were modeled as acting in parallel, independently of one another, rather than in a serial sequence.

2.1. Participants and Sample

The population consisted of undergraduate university students enrolled in Peruvian higher education institutions during the 2024–2025 academic period. A stratified non-probability convenience sample was used, considering the type of university (public/private), area of knowledge (social sciences, engineering, health sciences, humanities), and academic cycle (initial: 1–4, intermediate: 5–7, advanced: 8–10).

As shown in Table 1, the sample consisted of 5854 participants from 15 Peruvian universities located in Lima, Arequipa, Cusco, Trujillo, Piura, Chiclayo, Huancayo, and Iquitos. Universities were recruited through institutional agreements with academic coordinators at each institution, who facilitated access to students during scheduled class sessions and administered the instrument as part of the research team. Participation was entirely voluntary and anonymous, with no academic credit, grade improvement, or monetary compensation offered to participants. Data collection was conducted over ten weeks (October–December 2024), with 45% completing the questionnaire in person and 55% via Google Forms. The average response time was 43 min. Visual stimuli were presented in randomized order through a standardized platform.

Table 1. Sample distribution (n = 5854).

Variable	Category	n (%)
Type of university	Public	3215 (54.9%)
	Private	2639 (45.1%)
Field of study	Social sciences and communication	1930 (33.0%)
	Engineering and technology	1650 (28.2%)
	Health sciences	1250 (21.4%)
	Humanities and education	1024 (17.5%)

Table 1. *Cont.*

Variable	Category	n (%)
Academic cycle	Beginner (1–4)	2108 (36.0%)
	Intermediate (5–7)	2265 (38.7%)
	Advanced (8–10)	1481 (25.3%)
Gender	Female	3162 (54.0%)
	Male	2569 (43.9%)
	Other/Prefer not to say	123 (2.1%)
Age	Age range: 18–29 years old	M = 21.8, SD = 2.6

Note. *n* = sample size; % = percentage within category.

2.2. Operationalization of Variables

As shown in Table 2, the conceptual model integrated the following variables: News credibility assessment ability (*Y*, dependent variable), AI literacy (*X*, independent variable), Awareness of political deepfakes (*M1*, mediator 1), Political news verification strategies (*M2*, mediator 2), Type of university (*W*, moderator), and control variables (previous exposure to AI, digital media consumption, field of study, gender, and academic cycle).

Table 2. Operationalization of variables.

Category	Variable	Description
Dependent (<i>Y</i>)	News credibility assessment ability	Accuracy in identifying authentic and AI-generated political news content.
Independent (<i>X</i>)	AI literacy	Knowledge and critical evaluation of AI-generated content.
Mediator (<i>M1</i>)	Awareness of political deepfakes	Knowledge of political deepfakes and their risks to news credibility.
Mediator (<i>M2</i>)	Political news verification strategies	Strategies used to verify political audiovisual content.
Moderator (<i>W</i>)	Type of university	Public or private university affiliation.
Control (<i>VC1</i>)	Previous exposure to AI	Prior training or experience with artificial intelligence.
Control (<i>VC2</i>)	Digital media consumption	Frequency of political content consumption on digital platforms.
Control (<i>VC3</i>)	Sociodemographics	Field of study, gender, and academic cycle.

Note. *Y* = dependent variable; *X* = independent variable; *M* = mediator; *W* = moderator; *VC* = control variable.

2.3. Instrument

The instrument comprised five scales and a visual identification test, reviewed through expert judgment (Aiken's $V > 0.80$). AI literacy (*X*) included 8 items, awareness of political deepfakes (*M1*) 7 items, political news verification strategies (*M2*) 6 items, previous exposure to AI (*VC1*) 4 items, and digital media consumption (*VC2*) 4 items. Type of university (*W*) was classified as public or private. News credibility assessment ability (*Y*) was measured using the political and news-related visual stimuli, while non-political stimuli were retained for complementary analyses. The complete instrument is provided in Appendix A.

As shown in Table 3, all variables had factor loadings greater than 0.70, AVE values greater than 0.50, and Cronbach's alpha coefficients greater than 0.80, confirming the convergent validity and reliability of the instrument.

Table 3. Confirmatory factor analysis of the instrument.

Variable	Items	Factor Loadings	AVE	α	M	SD
AI literacy	8	0.762–0.831	0.628	0.893	3.48	0.73
Awareness of political deepfakes	7	0.748–0.819	0.564	0.879	3.72	0.68
Verification strategies	6	0.733–0.801	0.553	0.857	3.31	0.76
Prior exposure to AI	4	0.724–0.783	0.578	0.832	3.41	0.77

Note. M = Mean; SD = Standard Deviation (5-point Likert scale); AVE = Average Variance Extracted; α = Cronbach's Alpha. All factor loadings > 0.70.

2.4. Visual Stimuli and Procedure

The complete questionnaire, comprising 45 items, is provided in Appendix A. The first section assessed AI literacy, awareness of deepfakes, verification strategies, and prior exposure using a 5-point Likert scale. The second section presented a visual identification test with 20 stimuli (10 authentic and 10 deepfakes of Peruvian and international public figures), combining manipulated content retrieved from Peruvian social media financial fraud campaigns with synthetic content generated during instrument design in mid-2024. Stimuli were selected for recognizability, variability in manipulation sophistication, and representation of political and non-political contexts, and were validated by five specialists (Aiken's $V > 0.80$). A pilot test with 250 students confirmed adequate response variance across stimuli, ruling out ceiling or floor effects. For the present study, only the political and news-related stimuli were used to calculate the news credibility assessment ability score (Y): correct classifications were summed and rescaled to a 0–8 range to match the metric of the Likert-based predictor scales. All predictors were standardized prior to analysis, so reported beta coefficients reflect standardized effects comparable in relative magnitude across variables.

Figure 1 presents comparative examples of the stimuli used in the study, showing authentic images alongside their AI-generated counterparts. The images with dotted borders correspond to AI-generated content, while the images without borders correspond to authentic content. In each pair, the visual similarity between the real and synthetic content can be observed, illustrating the level of sophistication that makes it difficult for users to identify deepfakes.

Figure 2 presents highly sophisticated deepfakes that circulated on Peruvian social media as part of financial scam campaigns. This content impersonated public figures through facial manipulation with lip synchronization to promote fraudulent investment schemes.

Given that data were collected through two different modes, an additional analysis was conducted to assess whether the collection mode was associated with identification ability. A point-biserial correlation between collection mode (0 = online, 1 = in-person) and identification ability scores was non-significant ($r = 0.03$, $p = 0.21$), indicating that the two formats produced comparable results and that mode of administration did not introduce systematic bias into deepfake identification scores.



Figure 1. Comparison of authentic stimuli and AI-generated images. (A) shows Candidate 1; (B) shows Candidate 2; (C) shows Candidate 3; (D) shows Candidate 4; (E) shows Candidate 5; and (F) shows Candidate 6, each juxtaposing an authentic frame with its AI-generated counterpart. Note. Images with dotted borders correspond to AI-generated content. Images without borders correspond to authentic content.



Figure 2. Examples of highly sophisticated deepfakes used in the study. (A,D,E) show fabricated government announcements; (B) shows a fraudulent investment platform; (C) shows a fake investment campaign; and (F) shows a fabricated political debate segment. Note. Faces have been blurred to protect the identity of the individuals depicted. Content originally circulated on Peruvian social media platforms as part of financial fraud campaigns.

2.5. Data Analysis

Data analysis was performed using RStudio v.2024.04 and jamovi v.2.5 with PROCESS v4.2. A parallel mediation model (Model 4) evaluated the indirect effects of AI literacy on news credibility assessment ability through awareness of political deepfakes and political news verification strategies. A conditional model (Model 5) evaluated whether the type of

university moderated the direct relationship between AI literacy and the outcome. Previous exposure to AI, digital media consumption, field of study, gender, and academic cycle were included as control variables. Statistical significance was set at $\alpha = 0.05$, with 5000 bootstrap samples for indirect effects.

3. Results

To analyze the relationships between variables, mediation and moderation models were applied using the PROCESS module in R, allowing us to evaluate the effects of AI literacy on the ability to identify deepfakes, considering the impact of mediating and moderating variables. Additionally, multiple linear regression was performed to quantify the influence of each predictor on the dependent variable, based on the variables operationalized in Table 2 ($n = 5854$).

As shown in Figure 3, the model examines the relationship between AI literacy and news credibility assessment ability through two parallel mediators, with type of university as the moderator and previous exposure to AI, digital media consumption, field of study, gender, and academic cycle as control variables.

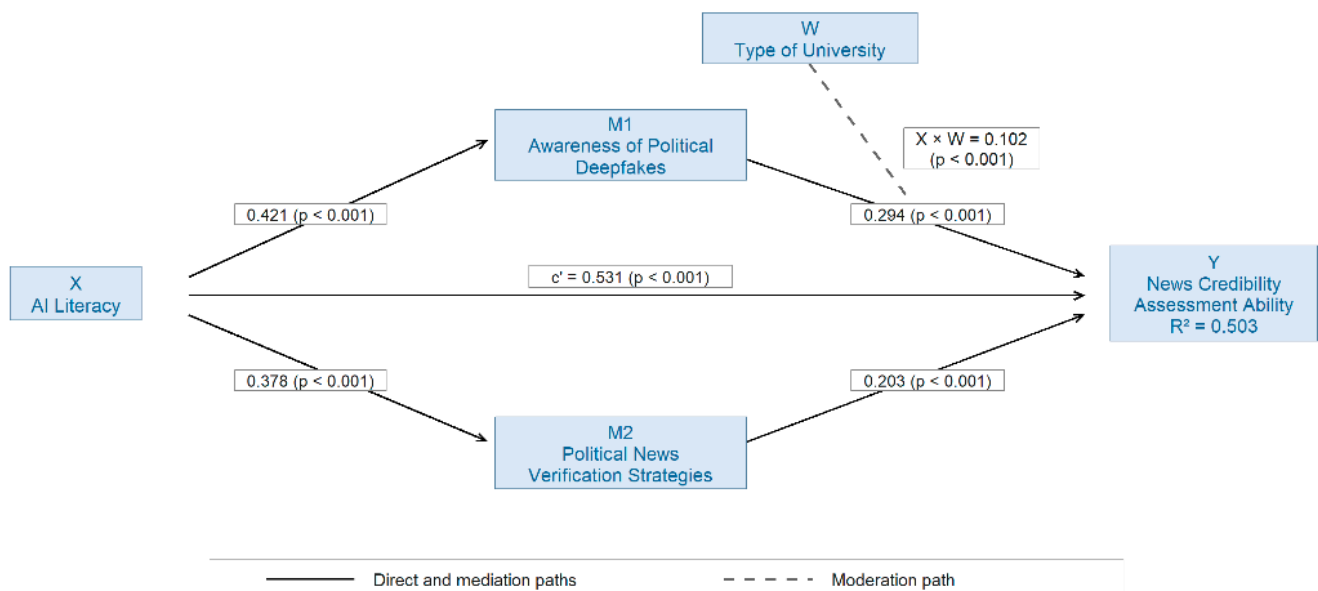


Figure 3. Conceptual and statistical model of mediation and moderation. Note. Solid black lines represent direct and mediation paths. Dashed lines represent the moderation of the direct effect (c'). Values shown as β (p -value).

3.1. Relationship Between the Variables in the Model

This model establishes the direct influence of AI literacy on mediating variables, namely awareness of deepfakes and verification strategies. The assumed causal structure, in which AI literacy precedes deepfake awareness and verification strategies, is theoretically grounded in the competency development literature: structured knowledge about AI systems is a prerequisite for meaningful risk awareness, as individuals must first understand what synthetic content is and how it is produced before they can reliably recognize it or apply systematic verification procedures (Bhuvaneswari et al., 2025; Huang & Hu, 2025). Experimental studies support this ordering by demonstrating that literacy interventions that build AI knowledge produce downstream improvements in both awareness and detection behavior, whereas awareness alone without foundational literacy training does not reliably improve identification performance (Zhao et al., 2025; Tassone et al., 2024). The relationship between AI literacy and identification ability is therefore both direct and

indirect, with part of the effect channeled through the mediating variables, a structure consistent with moderated mediation designs applied in prior digital literacy research (Lao et al., 2025).

In terms of moderation, the type of university conditions the direct relationship between AI literacy and news credibility assessment ability. The mediators, awareness of political deepfakes and political news verification strategies, were estimated through a parallel mediation model (PROCESS Model 4), while the moderating role of the type of university was evaluated separately using PROCESS Model 5. The same set of sociodemographic and behavioral control variables was included in both models.

3.2. Statistical Indicators and Mediation Analysis of the Model

As shown in Table 4, the model results reflect a moderate to high correlation ($R = 0.709$) between the variables, while the coefficient of determination $R^2 = 0.503$ indicates that 50.3% of the variability in Y is explained by the variables included in the model. The MSE (2.684) indicates a low level of error in the model's prediction. In addition, the F statistic (657.17, $p < 0.001$) indicates that the model is significant as a whole. The mediation analysis in the same table shows that both indirect effects exclude zero in their bootstrap confidence intervals, supporting mediation. The indirect effect through awareness of political deepfakes (M1) was 0.124 (95% CI [0.089, 0.161]) and through verification strategies (M2) was 0.077 (95% CI [0.051, 0.106]). The total indirect effect was 0.201 (95% CI [0.156, 0.247]), indicating that approximately 27.5% of the total effect of AI literacy on news credibility assessment ability is channeled through the two mediators.

Table 4. Statistical indicators and mediation analysis of the model (PROCESS Model 4; Dependent variable: News credibility assessment ability, Y).

R	R ²	MSE	F	df1	df2	p
0.709	0.503	2.684	657.17	9	5844	<0.001
Mediation Analysis: Indirect Effects (Bootstrap $n = 5000$)						
Path	Path-a	Path-b	Indirect effect	Boot SE	Boot LLCI	Boot ULCI
X → M1 → Y	0.421	0.294	0.124	0.018	0.089	0.161
X → M2 → Y	0.378	0.203	0.077	0.014	0.051	0.106
Total indirect effect			0.201	0.023	0.156	0.247
Direct effect (c')			0.531	0.020	0.492	0.570
Total effect (c)			0.732	0.028	0.677	0.787

Note. R = multiple correlation coefficient; R^2 = coefficient of determination; MSE = mean squared error; F = Fisher statistic; df1/df2 = degrees of freedom; CI = confidence interval based on 5000 bootstrap resamples (indirect effects are significant when CI excludes zero). X = AI literacy; M1 = Awareness of political deepfakes; M2 = Verification strategies; Y = News credibility assessment ability. Model estimated via PROCESS v4.2 (Model 4).

3.3. Linear Regression Analysis

As shown in Table 5, the linear regression analysis establishes the direct effects of each predictor on news credibility assessment ability (Y). AI literacy (X) shows the strongest direct effect ($\beta = 0.531$, $p < 0.001$), consistent with structured AI training being the strongest individual correlate of detection performance. Awareness of political deepfakes (M1) ($\beta = 0.294$, $p < 0.001$) and verification strategies (M2) ($\beta = 0.203$, $p < 0.001$) also show significant direct associations with Y. These coefficients reflect direct path-b relationships between each variable and the outcome, and do not constitute evidence of mediation on their own. The mediation paths, including path-a coefficients and indirect effect indices, are reported in the PROCESS mediation models in the following section.

Table 5. Direct effects and control variables associated with news credibility assessment ability (Y) (PROCESS Model 4).

Predictor	Coeff	SE	t	p	LLCI	ULCI
Intercept	3.550	0.021	173.120	<0.001	3.510	3.590
AI literacy (X)	0.531	0.014	39.248	<0.001	0.504	0.558
Awareness of political deepfakes (M1)	0.294	0.013	22.718	<0.001	0.269	0.319
Political news verification strategies (M2)	0.203	0.013	16.010	<0.001	0.178	0.228
Previous exposure to AI (VC1)	0.162	0.012	13.602	<0.001	0.139	0.185
Digital media consumption (VC2)	−0.092	0.011	−8.037	<0.001	−0.114	−0.070
Field: Engineering (VC3)	0.141	0.025	5.554	<0.001	0.091	0.191
Male gender (VC3)	0.071	0.023	3.091	0.002	0.026	0.116
Type of university: Private (W)	0.086	0.023	3.760	<0.001	0.041	0.131
Advanced cycle (VC3)	0.062	0.026	2.354	0.019	0.010	0.114

Note. Coeff = unstandardized coefficient; SE = standard error; LLCI/ULCI = lower/upper limits of the 95% confidence interval. X = independent variable; M = mediator; VC = control variable; W = moderator variable.

Additionally, two control variables theoretically relevant to digital literacy also show direct associations with news credibility assessment ability in this regression model: previous exposure to AI ($\beta = 0.162$, $p < 0.001$) shows a positive direct association with identification ability, while digital media consumption ($\beta = -0.092$, $p < 0.001$) has a negative effect, indicating that greater exposure without critical training reduces detection ability. The remaining control variables also show significant direct effects: belonging to the engineering field ($\beta = 0.141$, $p < 0.001$), male gender ($\beta = 0.071$, $p < 0.001$), and advanced cycle ($\beta = 0.062$, $p < 0.001$) are associated with greater identification ability. Type of university ($\beta = 0.086$, $p < 0.001$) also shows a significant direct association here, though its role as a moderator of the AI literacy–identification relationship is examined separately in Section 3.4. All predictors were statistically significant ($p < 0.001$), validating the robustness of the model. Regression diagnostics confirmed the adequacy of the model assumptions. Variance Inflation Factors (VIF) for all predictors ranged from 1.08 to 1.97, indicating no multicollinearity concerns. Residual plots confirmed approximate homoscedasticity across the range of fitted values. Full regression diagnostics are reported in Appendix B.

Given the large sample size ($n = 5854$), all predictors reached statistical significance at $p < 0.001$, which is expected even for associations of modest practical magnitude and should not, on its own, be read as evidence of practical relevance. To evaluate substantive significance alongside statistical significance, the standardized coefficients in Table 5 can be compared in relative terms: AI literacy ($\beta = 0.531$) has an effect roughly 3.7 times larger than academic cycle ($\beta = 0.062$) and roughly 6 times larger than digital media consumption ($\beta = -0.092$), indicating that, among the predictors examined, AI literacy is not only statistically significant but also the practically dominant correlate of identification ability. By contrast, sociodemographic and institutional control variables (gender, academic cycle), while statistically significant, show standardized coefficients below 0.10, suggesting comparatively modest practical relevance on their own, even though the underlying disparities they capture may still be pedagogically and institutionally meaningful. Following the logic of a smallest effect size of interest (SESIOI), future research and applied interpretation of these findings should treat statistical significance as a necessary but not sufficient criterion, and should consider pre-registering minimum effect sizes of practical interest before treating the control variables as targets for intervention.

3.4. Moderation Analysis

As shown in Table 6, AI literacy was positively associated with news credibility assessment ability among students from both public and private universities. The interaction between AI literacy and type of university was statistically significant ($\beta = 0.102$, $p < 0.001$), indicating that the strength of this relationship differed according to institutional affiliation. The conditional effect was stronger among students from private universities ($\beta = 0.588$, 95% CI [0.552, 0.625]) than among students from public universities ($\beta = 0.486$, 95% CI [0.453, 0.519]). These results suggest that institutional context moderates the relationship between AI literacy and the ability to assess the credibility of political news.

Table 6. Conditional direct effects of AI literacy on news credibility assessment ability (Y) according to moderating variables (PROCESS Model 5).

Effect	Coeff	SE	t	p	LLCI	ULCI
AI literacy (X)	0.486	0.017	28.895	<0.001	0.453	0.519
Type of university (W)	0.086	0.023	3.760	<0.001	0.041	0.131
AI literacy $\times$ Type of university	0.102	0.023	4.461	<0.001	0.057	0.147
Conditional effect: Public university	0.486	0.017	28.895	<0.001	0.453	0.519
Conditional effect: Private university	0.588	0.019	31.577	<0.001	0.552	0.625

Note. Public university was coded as 0 and private university as 1. The interaction term represents the moderating effect of the type of university on the relationship between AI literacy and news credibility assessment ability. X = independent variable (AI literacy); W = moderator variable (type of university); SE = standard error; LLCI/ULCI = lower and upper limits of the 95% confidence interval. Previous exposure to AI, digital media consumption, field of study, gender, and academic cycle were included as control variables.

3.5. Graphical Interpretation by University Type

As shown in Figure 4, the association between AI literacy (X) and news credibility assessment ability (Y) is positive for students from both public and private universities, but the strength of this association differs by institutional type. Students from private universities show a steeper slope ($\beta = 0.588$) than students from public universities ($\beta = 0.486$), indicating that the practical benefit of AI literacy on news credibility assessment ability is somewhat stronger among private university students. These patterns are consistent with the type of university acting as a moderator of this association, although the cross-sectional design does not allow causal claims about this moderating role.

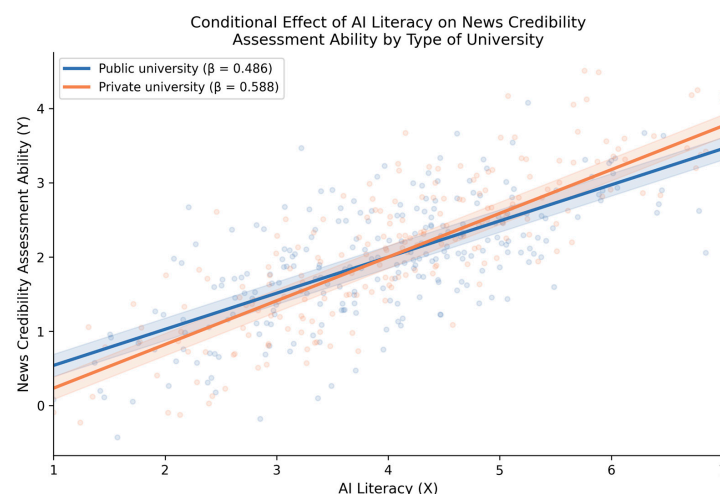


Figure 4. Effect of AI literacy on news credibility assessment ability by type of university. Note. Lines represent conditional regression slopes of AI literacy on news credibility assessment ability for public ($\beta = 0.486$) and private ($\beta = 0.588$) universities. Shaded bands represent 95% confidence intervals. All paths $p < 0.001$.

4. Discussion

The results provide strong empirical support for the proposed mediation-moderation model, showing that AI literacy is the strongest individual correlate of deepfake identification ability among Peruvian university students, both directly and through the mediating roles of deepfake awareness and verification strategies. The model explained a substantial portion of the variance in identification ability, which positions AI literacy as a structural competency rather than a peripheral skill in digital media education.

The first finding indicates that AI literacy is the strongest individual correlate of deepfake identification ability, a result that carries direct implications for how universities design curricula in the context of political disinformation. The practical implication is not merely that AI training helps students spot fake content, but that without it, even digitally active students remain structurally vulnerable to manipulation regardless of how frequently they consume online media. This aligns with the broader argument that AI literacy functions as a threshold competency: below a certain level of foundational understanding, exposure to digital content does not build critical discernment but instead reinforces habitual and uncritical information processing (Celik, 2023; Labuz, 2025). The experimental evidence further supports this interpretation, showing that structured media literacy interventions consistently outperform passive awareness campaigns in producing reliable detection improvements (Hynek et al., 2025).

The mediating roles of deepfake awareness and verification strategies reveal the mechanism through which AI literacy translates into detection performance, and this mechanism has important practical consequences for news credibility. Deepfake awareness mediates the relationship more strongly than verification strategies, suggesting that conceptual understanding of how synthetic content works is a more powerful intermediary than the behavioral application of checking procedures. This finding challenges the assumption that teaching verification techniques is sufficient: without the conceptual foundation that awareness provides, verification strategies may be applied inconsistently or abandoned when content appears credible at first glance. The implication for journalism education is significant: young people generally express skepticism toward digital content but consistently fall back on intuitive heuristics when confronted with sophisticated audiovisual material, meaning that skepticism without structured competency does not protect against politically motivated deepfakes (Sánchez-Esparza et al., 2024). Restoring trust in news therefore requires building both conceptual awareness and systematic behavioral habits, since neither alone is sufficient (Battista, 2024; Momeni, 2025).

The moderating findings reveal a structural equity problem that extends beyond individual competency gaps. Type of university significantly moderates the relationship between AI literacy and news credibility assessment ability, with private university students deriving a stronger practical benefit from AI literacy ($\beta = 0.588$) than their public university counterparts ($\beta = 0.486$). This gap is unlikely to reflect differences in the training itself; rather, it points to unequal complementary conditions, such as institutional infrastructure, curricular design, and extracurricular exposure to AI tools, that shape how effectively literacy translates into detection competence. Students from public universities, who make up the majority of this sample, are therefore left in a comparatively weaker position even when they receive equivalent AI literacy training, suggesting that existing inequalities in access to technology and institutional resources are being reproduced within higher education (Gökçe & Nacaroglu, 2025; Khan et al., 2025). Beyond this moderating effect, digital media consumption showed a negative direct association with detection ability, indicating that high-frequency media consumers without critical training show reduced detection capacity; exposure without structure appears to degrade discernment rather than build it (Kaewsawad & Duda, 2025).

The influence of sociodemographic and institutional variables on identification ability points to a deepening AI divide within Peruvian higher education that carries consequences for news credibility at a societal level. Engineering students demonstrate systematically higher detection capabilities as a direct effect of the field of study, while the type of university operates as a moderator that amplifies or dampens the returns to AI literacy itself, reflecting differential institutional capacity to translate training into competence rather than differences in innate ability. If left unaddressed, this structural gap means that the students most likely to enter journalism, education, communications, and public administration, fields where the ability to evaluate the authenticity of political audiovisual content is most consequential, are precisely those receiving training that yields comparatively smaller returns in detection competence when delivered under public university conditions (Özüdoğru & Durak, 2025). The policy implication is direct: AI literacy programs cannot be designed as elective or discipline-specific offerings without reproducing and widening existing inequalities. Strengthening the institutional conditions, infrastructure, faculty preparation, and applied practice opportunities, under which AI literacy training is delivered at public universities is necessary to prevent a scenario in which the capacity to identify political deepfakes becomes a privilege of private institutions rather than a shared democratic competency (Lao et al., 2025; Zhao et al., 2025).

5. Conclusions

The findings of this study confirm that AI literacy is an effective mechanism for improving the ability of Peruvian university students to identify deepfakes, especially when complemented by the development of awareness about deepfakes and systematic verification strategies. It was found that knowledge about AI and the application of verification strategies directly influence students' performance when faced with synthetic content (Huang & Hu, 2025; Zhao et al., 2025).

Likewise, the results indicate that the type of university significantly moderates this relationship, with private university students showing a greater practical benefit from AI literacy than their public university counterparts (Kazanidis & Pellas, 2024; Ofem et al., 2024). In this sense, educational strategies need to consider an integrated approach where AI training is combined with critical thinking and verification skills, while also addressing the institutional gaps that shape its effectiveness.

The influence of the area of knowledge and type of university on the adoption of digital media literacy skills is noteworthy. It was confirmed that engineering students and students at private universities have greater access to AI training and adopt more effective verification strategies (Yang, 2024). This highlights the need to implement AI literacy programs across all areas of knowledge, with particular attention to public universities and humanities and social science programs, where the risk of a widening AI divide is most acute if targeted interventions are not implemented.

This study contributes to the debate on digital media literacy in Latin American contexts, emphasizing that effective training programs must integrate technical understanding of AI, awareness of the risks of deepfakes, and training in systematic verification strategies. The combination of these factors ensures a safer digital environment, strengthens news credibility in politically polarized contexts, and reduces the vulnerability of Peruvian university students to misinformation generated by artificial intelligence. For journalism and media institutions operating in Peru, these findings point to the urgency of supporting AI literacy initiatives that protect the integrity of public information and democratic discourse.

Limitations

This study has four limitations. First, the cross-sectional design establishes correlational associations rather than causal relationships; experimental designs with pre- and post-intervention measurement are needed to confirm causality. Second, the identification task may inflate false positive rates, as asking participants to classify stimuli as authentic or AI-generated may induce suspicion regardless of actual content. Third, the mixed data collection mode may have introduced variation in responses, since deepfakes are typically encountered in digital rather than printed formats. Fourth, survey-based identification performance may overestimate real-world detection capacity, as naturalistic conditions involve higher information volume, time pressure, and emotional engagement than a controlled research setting. In an era of information abundance, the sheer volume of audiovisual content circulating across digital platforms makes systematic verification of every piece of content informationally and cognitively unfeasible for most citizens, which limits the practical transferability of the detection skills measured in this study.

Author Contributions: Conceptualization, R.B.A.C., A.A.M.C. and M.O.R.S.; methodology, R.B.A.C., A.A.M.C., L.D.Z.C. and A.L.F.; software, A.E.D.S.; validation, A.A.M.C., L.D.Z.C., A.L.F., A.V.M.G. and A.E.D.S.; formal analysis, L.D.Z.C., A.L.F. and A.E.D.S.; investigation, A.A.M.C., J.E.F.C., A.V.M.G., J.L.C.C. and M.O.R.S.; resources, A.V.M.G. and J.L.C.C.; data curation, L.D.Z.C., J.E.F.C., A.V.M.G. and J.L.C.C.; writing—original draft preparation, J.E.F.C. and M.O.R.S.; writing—review and editing, R.B.A.C., A.A.M.C., L.D.Z.C., A.L.F., J.L.C.C. and M.O.R.S.; visualization, J.E.F.C. and A.E.D.S.; supervision, R.B.A.C.; project administration, R.B.A.C. Funding acquisition was omitted as this research received no external funding. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived for this study because the research involved exclusively adult university students (aged 18–29) who were not in a position of vulnerability, participated voluntarily, and provided informed consent prior to data collection. The study used anonymous self-reported survey data and a visual identification test without any clinical intervention, physical risk, or collection of sensitive personal information. All procedures were conducted in accordance with the ethical principles established in the Declaration of Helsinki (2013 revision) and in compliance with Peru's Personal Data Protection Law (Ley N° 29733) and its Regulations (D.S. N° 003-2013-JUS), which guarantee the confidentiality and anonymity of participants' information.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy restrictions related to the participants' institutional affiliation and the confidentiality agreement established during the data collection process.

Acknowledgments: The authors express their gratitude to the 15 Peruvian universities and the 5854 students who participated in this study, as well as to the academic authorities and faculty members who facilitated access to the study participants.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Complete Measurement Instrument

Please mark with an X (or circle) the number that best describes your situation or degree of agreement with each statement. There are no right or wrong answers. The information provided is strictly confidential and will be used for academic purposes only.

1 = Strongly disagree 2 = Disagree 3 = Neither agree nor disagree 4 = Agree
5 = Strongly agree.

Appendix A.1. Self-Report Scales

Table A1. Complete Measurement Instrument (Self-Report Scales and Visual Identification Test).

No.	ITEM	1	2	3	4	5
VARIABLE 1: AI LITERACY (X)						
Dimension 1: Technical Understanding						
01	I understand how artificial intelligence systems process and generate information.					
02	I can distinguish between content produced by humans and content generated by AI systems.					
03	I have sufficient technical knowledge to identify when an image or video may have been manipulated using AI.					
Dimension 2: Critical Thinking about AI						
04	I am aware of the potential biases embedded in AI-generated outputs.					
05	I can critically evaluate the reliability of information produced by AI tools.					
06	I can explain to others how deepfake generation technology works.					
Dimension 3: Ethical Evaluation						
07	I understand the ethical implications of using AI-generated content in public communication.					
08	I regularly update my knowledge about new developments in artificial intelligence.					
VARIABLE 2: AWARENESS OF POLITICAL DEEPPAKES (M1)						
Dimension 1: Conceptual Knowledge						
09	I know what deepfakes are and how they are produced.					
10	I am aware that deepfakes can be used to manipulate public opinion and political discourse.					
11	I perceive deepfakes as a significant threat to the credibility of news and journalism.					
Dimension 2: Risk Perception						
12	I believe that AI-generated audiovisual content poses risks to democratic processes.					
13	I trust digital media content less because of the existence of deepfake technology.					
Dimension 3: Trust in Media						
14	I am concerned about my own ability to detect deepfakes in my daily media consumption.					
15	I believe that most people around me are vulnerable to being deceived by deepfakes.					
VARIABLE 3: VERIFICATION STRATEGIES (M2)						
Dimension 1: Source Verification						

Table A1. Cont.

No.	ITEM	1	2	3	4	5
16	When I encounter suspicious content online, I verify it using trusted news sources before sharing it.					
17	I use reverse image search tools to check whether a photo has been manipulated.					
Dimension 2: Technical Analysis						
18	I consult fact-checking platforms when I am unsure about the authenticity of digital content.					
19	I analyze visual details such as lighting, shadows, and facial features to detect potential manipulations.					
Dimension 3: Consultation with Experts						
20	I seek expert opinion or specialized tools when I suspect a video or image may be AI-generated.					
21	I apply systematic verification procedures before accepting audiovisual content as authentic.					
VARIABLE 4: PRIOR EXPOSURE TO AI (VC1)						
Dimension 1: Formal and Informal Training						
22	I have received formal academic training on artificial intelligence or machine learning.					
23	I have participated in workshops, courses, or seminars related to AI tools or digital media literacy.					
24	I have independently studied topics related to AI through online platforms or self-directed learning.					
25	I have prior experience using AI-based tools for content creation, analysis, or detection.					
VARIABLE 5: DIGITAL MEDIA CONSUMPTION (VC2)						
Items 26–27: 1 = Less than 1 h/day; 2 = 1–2 h; 3 = 3–4 h; 4 = 5–6 h; 5 = More than 6 h/day						
26	On average, how many hours per day do you consume content on social media platforms?					
27	On average, how many hours per day do you consume news or informational content through digital platforms?					
Items 28–29: 1 = Never; 2 = Rarely; 3 = Sometimes; 4 = Often; 5 = Always						
28	How frequently do you share digital content (images, videos, articles) without verifying its authenticity?					
29	How frequently do you encounter content that you suspect may be AI-generated or manipulated?					
SOCIODEMOGRAPHIC AND INSTITUTIONAL INFORMATION						
30	30 Field of study (VC3): (1) Social sciences and communication (2) Engineering and technology (3) Health sciences (4) Humanities and education					
31	Gender (VC3): (1) Female (2) Male (3) Other/Prefer not to say					
32	Type of university (W): (1) Public (2) Private					
33	Academic cycle (VC3): (1) Beginner (cycles 1–4) (2) Intermediate (cycles 5–7) (3) Advanced (cycles 8–10)					
34	Age (VC3): _____					

Appendix A.2. Visual Identification Test: 20 Stimuli

Adapted from [Huang and Hu \(2025\)](#)

Instructions: You will be shown 20 images one at a time. For each image, read the question below, mark your response with an X, and briefly explain the reason for your decision.

Question: Is the image you just saw authentic, or was it generated/manipulated by artificial intelligence?

Stimulus 1

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 2

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 3

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 4

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 5

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 6

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 7

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 8

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 9

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 10

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 11

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 12

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 13

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 14

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 15

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 16

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 17

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 18

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 19

☐ The content is authentic

☐ The content was generated or manipulated by AI

Reason for your answer:

Stimulus 20
[] The content is authentic
[] The content was generated or manipulated by AI
Reason for your answer:

Note. Stimuli 1–10 correspond to authentic political and news-related content, while stimuli 11–20 correspond to AI-generated or manipulated political and news-related content. Stimulus order was randomized for each participant. The news credibility assessment ability score (Y) was calculated by dividing the number of correct classifications by 20 and multiplying by 8, yielding a score range of 0 to 8 points.

Appendix B. Regression Diagnostics

The following tables report regression diagnostics for the main mediation-moderation model. Results confirm that the standard assumptions of multiple linear regression are met.

Table A2. Variance Inflation Factors (VIF) for All Predictors. VIF values below 5.0 indicate no multicollinearity. Tolerance values above 0.20 confirm predictor independence.

Predictor	VIF	Tolerance
AI literacy (X)	1.43	0.699
Awareness of political deepfakes (M1)	1.67	0.599
Verification strategies (M2)	1.58	0.633
Prior exposure to AI	1.38	0.725
Digital media consumption	1.21	0.826
Field: Engineering (VC)	1.19	0.840
Male gender (VC)	1.08	0.926
Private university (VC)	1.14	0.877
Advanced cycle (VC)	1.27	0.787

Note. All VIF values are below the conventional threshold of 5.0, indicating no multicollinearity concerns. Tolerance values above 0.20 confirm predictor independence.

Table A3. Homoscedasticity Test (Breusch-Pagan). A non-significant result ($p > 0.05$) confirms that the homoscedasticity assumption is met.

Test	Statistic	df	p
Breusch-Pagan	11.43	9	0.248

Note. The non-significant Breusch-Pagan test ($p = 0.248$) confirms that the homoscedasticity assumption is met. Residual plots confirmed approximately uniform variance across fitted values.

References

Barthes, R. (1981). *Camera lucida: Reflections on photography* (R. Howard, Trans.). Hill and Wang. (Original work published 1980).

Battista, D. (2024). Political communication in the age of artificial intelligence: An overview of deepfakes and their implications. *Society Register*, 8(2), 7–24. [CrossRef]

Bender, S. M. (2024). Awareness of artificial intelligence as an essential digital literacy: ChatGPT and Gen-AI in the classroom. *Changing English: Studies in Culture and Education*, 31(2), 161–174. [CrossRef]

Bhandari, R., & Bhandari, S. (2025). Artificial intelligence: Understanding deepfakes. *EDPACS*, 70(9), 21–31. [CrossRef]

Bhuvaneswari, T., Chandra Guru Sekar, R., Shiva Raama Krishna, K., Nishaal, S. R., & Arun Kumar, M. (2025). Detection of deepfake videos in social networks. *Signal, Image and Video Processing*, 19(13), 1160. [CrossRef]

Bircan, M. A., Şeref, İ., & Nacaroglu, O. (2025). The effect of STEM themed story writing training with artificial intelligence tools on the digital literacy and 21st century skills of preservice teachers. *Education and Information Technologies*, 30(15), 21479–21498. [CrossRef]

Celik, I. (2023). Exploring the determinants of artificial intelligence (AI) literacy: Digital divide, computational thinking, cognitive absorption. *Telematics and Informatics*, 83, 102026. [CrossRef]

Duc Truong, K. C. (2024). Reputation (not Taylor’s version): Regulating artificial intelligence hallucinated deepfakes of public figures. *Journal of Law, Technology and Policy*, 2024(2), 449–476.

- Garriaga, M., Ruiz-Incertis, R., & Magallón-Rosa, R. (2024). Inteligencia artificial, desinformación y propuestas de alfabetización en torno a los deepfakes [Artificial intelligence, disinformation and media literacy proposals around deepfakes]. *Observatorio*, 18, 175–194. [CrossRef]
- Gökçe, H., & Nacaroglu, O. (2025). The effect of the use of artificial intelligence tools in science education on secondary school students' 21st century skills competency perceptions and digital literacy. *Education and Information Technologies*, 31, 1059–1077. [CrossRef]
- Gronseth, S. L., Itani, A., Seastrand, K., Beech, B. M., Bruce, M. A., Solorio, T., & Kakadiaris, I. A. (2025). Exploring a digital educational escape room game to support artificial intelligence literacy and team building in an online environment. *Journal of Interactive Learning Research*, 36(1), 29–44. [CrossRef]
- Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior, and Social Networking*, 24(3), 149–152. [CrossRef] [PubMed]
- Huang, G., & Hu, B. (2025). “A warning is not enough. Teach me how to spot deepfakes”: Testing media literacy interventions for combating deepfakes. *Science Communication*. Advance online publication. [CrossRef]
- Hynek, N., Gavurová, B., & Kubák, M. (2025). Risks and benefits of artificial intelligence deepfakes: Systematic review and comparison of public attitudes in seven European countries. *Journal of Innovation and Knowledge*, 10(5), 100782. [CrossRef]
- Kaewsawad, V., & Duda, J. (2025). Assessing artificial intelligence (AI) literacy and readiness in Thailand's workforce: Challenges and opportunities for digital transformation. *Science and Technology Asia*, 30(3), 209–231.
- Kazanidis, I. K., & Pellas, N. (2024). Harnessing generative artificial intelligence for digital literacy innovation: A comparative study between early childhood education and computer science undergraduates. *AI*, 5(3), 1427–1445. [CrossRef]
- Khan, T., Nag, A. K., Basu, S., & Saluja, V. (2025). Artificial intelligence for development (AI4D) usage, learner autonomy, wellbeing and digital literacy in ODL. *Journal of Learning for Development*, 12(2), 330–346. [CrossRef]
- Kontos, I., Chrysanthopoulou, K., & Galanopoulos-Papavasileiou, I. (2025). Photojournalism in the age of deepfakes: The role of media literacy and ethical standards in restoring trust in visual reporting. *Journal of Communication and Media Studies*, 10(2), 1–22. [CrossRef]
- Lao, Y., Hirvonen, N., & Larsson, S. (2025). Everyday encounters with deepfakes: Young people's media and information literacy practices with AI-generated media. *Journal of Documentation*, 81(7), 216–235. [CrossRef]
- Lin, X., Xu, G., & Xiong, B. (2025). Artificial intelligence literacy, sustainability of digital learning and practice achievement: A study of vocational college students. *PLoS ONE*, 20(10), e0332175. [CrossRef] [PubMed]
- Łabuz, M. (2025). A teleological interpretation of the definition of deepfakes in the EU artificial intelligence act—A purpose-based approach to potential problems with the word “existing”. *Policy and Internet*, 17(1), e435. [CrossRef]
- Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys*, 54(1), 7. [CrossRef]
- Momeni, M. (2025). Artificial intelligence and political deepfakes: Shaping citizen perceptions through misinformation. *Journal of Creative Communications*, 20(1), 41–56. [CrossRef]
- Morales Gonzales, A. V., Zapata Manrique, C. R., Gamero Juarez, L. R., Cuentas Galindo, I., Campos Ascuña, L. M., Almanza Cabe, R. B., Saenz Colana, A. S., Arguedas-Catani, R. W., & Rodriguez-Saavedra, M. O. (2026). Normalization of corruption and evaluation of public services: Adaptive mechanisms and subnational heterogeneity in Peru. *Frontiers in Political Science*, 8, 1806986. [CrossRef]
- Muawanah, U., Marini, A., & Sarifah, I. (2024). The interconnection between digital literacy, artificial intelligence, and the use of E-learning applications in enhancing the sustainability of Regional Languages: Evidence from Indonesia. *Social Sciences and Humanities Open*, 10, 101169. [CrossRef]
- Musyaffi, A. M., Adha, M. A., Mukhibad, H., & Oli, M. C. (2024). Improving students' openness to artificial intelligence through risk awareness and digital literacy: Evidence from a developing country. *Social Sciences and Humanities Open*, 10, 101168. [CrossRef]
- Neyazi, T. A., Khai Ee, T., & Kuru, O. (2025). Campaign deepfakes and affective polarization: The role of artificial intelligence in campaigns in shaping voter attitudes. *Social Science Computer Review*. Advance online publication. [CrossRef]
- Obateru, O. T., Oyedokun-Alli, W. A., & Shittu, S. B. (2025). Unpacking the relationship among undergraduates' artificial intelligence anxiety, readiness, perceived competence, and digital literacy. *International Journal of Learning in Higher Education*, 33(1), 97–119. [CrossRef]
- Ofem, U. J., Asuquo, E. N., Akeke, M. N. G., Idung, J. U., Anake, P. M., Ajuluchukwu, E. N., Ene, E. I., Amanso, E. O. I., Edam-Agbor, I. B., Okute, A. L., Anyin, N. N., Orim, F. S., Ekpang, P. O., Atah, C. A., Okim, O. T., Orji, E. I., & Echu, A. E. (2024). Curriculum factors and sustainable artificial intelligence-driven classroom assessment. The mediating role of computer self-efficacy and digital literacy. *Journal of Applied Learning and Teaching*, 7(2), 206–222. [CrossRef]
- Okolie, C. (2023). Artificial intelligence-altered videos (deepfakes), image-based sexual abuse, and data privacy concerns. *Journal of International Women's Studies*, 25(2), 11.
- Özüdoğru, G., & Durak, H. Y. (2025). Conceptualizing pre-service teachers' artificial intelligence readiness and examining its relationship with various variables: The role of artificial intelligence literacy, digital citizenship, artificial intelligence-enhanced

- innovation and perceived threats from artificial intelligence [Special issue: “Artificial intelligence initiatives”]. *Information Development*, 41(3), 916–932. [\[CrossRef\]](#)
- Pedersen, K. T., Pepke, L., Stærmose, T., Papaioannou, M., Choudhary, G., & Dragoni, N. (2025). Deepfake-driven social engineering: Threats, detection techniques, and defensive strategies in corporate environments. *Journal of Cybersecurity and Privacy*, 5(2), 18. [\[CrossRef\]](#)
- Rodriguez-Saavedra, M. O., Cuentas Galindo, I., Campos Ascuña, L. M., Morales Gonzales, A. V., Chavez Cuarite, J. L., & Almanza Cabe, R. B. (2026a). AI literacy and deepfake detection competence in Peruvian higher education. *European Journal of Educational Research*, 15(3), 1009–1023. [\[CrossRef\]](#)
- Rodriguez-Saavedra, M. O., Donayre Prado, E. A., & Donayre Sarolli, A. E. (2026b). From getting out to vote: Digital political narratives and their effect on young people’s electoral decision-making. *VISUAL REVIEW. International Visual Culture Review Revista Internacional de Cultura Visual*, 18(1), 241–260. [\[CrossRef\]](#)
- Rodriguez Saavedra, M. O., Lopez, J. M. B., Nina, W. Q., Morales Gonzales, A. V., Galindo, I. C., Ascuña, L. M. C., Colana, A. S. S., Almanza Cabe, R. B., Tito, P. G. L., & Teran, S. V. L. (2026). Machine learning and generative AI in administrative processes in Peru: Administrative efficiency in the national public sector. *Informatics*, 13(3), 44. [\[CrossRef\]](#)
- Rodriguez-Saavedra, M. O., Morales Gonzales, A. V., Cuentas Galindo, I., Campos Ascuña, L. M., Mamani Lopez, J. W., Caira Mamani, R. N., Liendo Terán, S. V., & Almanza Cabe, R. B. (2026c). Political communication and digital participation in Arequipa, Peru: Social media consumption and media credibility. *Studies in Media and Communication*, 14(2), 284–298. [\[CrossRef\]](#)
- Saklaki, A., & Gardikiotis, A. (2024). Exploring Greek students’ attitudes toward artificial intelligence: Relationships with AI ethics, media, and digital literacy. *Societies*, 14(12), 248. [\[CrossRef\]](#)
- Sanchez-Acedo, A., Carbonell-Alcocer, A., Gertrudix, M., & Rubio-Tamayo, J. L. (2024). Retos de la alfabetización mediática e informacional en la ecología de la inteligencia artificial: Deepfakes y desinformación resumen [The challenges of media and information literacy in the artificial intelligence ecology: Deepfakes and misinformation; Special issue]. *Communication and Society*, 37(4), 223–239. [\[CrossRef\]](#)
- Sánchez-Esparza, M., Palella-Stracuzzi, S., & Fernández Fernández, Á. (2024). Impacto de la inteligencia artificial en RTVE: Verificación de vídeos falsos y deepfakes, generación de contenidos y nuevos perfiles profesionales [Impact of artificial intelligence on RTVE: Verification of fake videos and deep fakes, content generation, and new professional profiles]. *Communication and Society*, 37(2), 261–277. [\[CrossRef\]](#)
- Seker, O., Kwon, K., & Kocak, O. (2025). Exploring researchers’ artificial intelligence (AI) literacy: The mediating role of digital literacy and data literacy between 21st century skills and AI literacy. *Information Development*. Advance online publication. [\[CrossRef\]](#)
- Sharma, P., Kumar, M. P., & Sharma, H. K. (2025). A robust ensemble model for deepfake detection of GAN-generated images on social media. *Discover Computing*, 28(1), 41. [\[CrossRef\]](#)
- Simó-Soler, E. S. (2023). Retos jurídicos derivados de la inteligencia artificial generativa: Deepfakes y violencia contra las mujeres como supuesto de hecho [Legal challenges arising from Generative artificial intelligence: Deepfakes and violence against women as a factual scenario]. *InDret*, 2, 493–515. [\[CrossRef\]](#)
- Sohaib, M., & Tehseen, S. (2023). Forgery detection of low quality deepfake videos. *Neural Network World*, 33(2), 85–99. [\[CrossRef\]](#)
- Tassone, F., Maiano, L., & Amerini, I. (2024). Continuous fake media detection: Adapting deepfake detectors to new generative techniques. *Computer Vision and Image Understanding*, 249, 104143. [\[CrossRef\]](#)
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1), 1–13. [\[CrossRef\]](#)
- Vecchietti, G., Liyanaarachchi, G., & Viglia, G. (2025). Managing deepfakes with artificial intelligence: Introducing the business privacy calculus. *Journal of Business Research*, 186, 115010. [\[CrossRef\]](#)
- Vilca Corrales, M. P., Caira Mamani, R. N., Jara Ortega, C. E., Cuentas Galindo, I., Campos Ascuña, L. M., Morales Gonzales, A. V., Saenz Colana, A. S., Almanza Cabe, R. B., Aroquipa-Apaza, O., & Rodríguez-Saavedra, M. O. (2026). Digital public service quality and results-oriented governance in Peru. *Frontiers in Political Science*, 8, 1788065. [\[CrossRef\]](#)
- Wang, B., Zhang, C., Khamo, & Ye, S. (2025). Facing the era of generative artificial intelligence: Strategies of information and digital literacy in Chinese studies. *IFLA Journal*, 51(3), 787–802. [\[CrossRef\]](#)
- Wang, H., Liu, Z., & Wang, S. (2023). Exploiting complementary dynamic incoherence for deepfake video detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(8), 4027–4040. [\[CrossRef\]](#)
- Wang, T., Cheng, H., Chow, K. P., & Nie, L. (2023). Deep convolutional pooling transformer for deepfake detection. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(6), 179. [\[CrossRef\]](#)
- Wu, D., & Zhang, J. (2025). Generative artificial intelligence in secondary education: Applications and effects on students’ innovation skills and digital literacy. *PLoS ONE*, 20(5), e0323349. [\[CrossRef\]](#) [\[PubMed\]](#)
- Yang, Y. (2024). Influences of Digital literacy and moral sensitivity on artificial intelligence ethics awareness among nursing students. *Healthcare*, 12(21), 2172. [\[CrossRef\]](#) [\[PubMed\]](#)

- Zhao, X., Narasuman, S. A., & Ren, W. (2025). La alfabetización digital de los profesores de inglés como lengua extranjera y su relación con la autoeficacia y el bienestar en un contexto basado en la inteligencia artificial [EFL teachers' digital literacy and its relation to self-efficacy and well-being in artificial intelligence-based context; Special issue]. *Porta Linguarum*, 2025(13), 391–408. [[CrossRef](#)]
- Zuo, Z., Luo, Y., Yan, S., & Jiang, L. (2025). From perception to practice: Artificial intelligence as a pathway to enhancing digital literacy in higher education teaching. *Systems*, 13(8), 664. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.