



Article

Legal Literacy, AI Risk Perception, and Digital Safety in UAE Universities: Testing a Trust-Calibrated Vigilance Model

Abdessalam Salmi ^{1,*} , Mohamed Amin Mekheimer ^{2,*} and Kayaty Ashour ³

¹ College of Law, Ajman University, Ajman P.O. Box 346, United Arab Emirates

² Faculty of Education, Beni-Suef University, Beni Suef 62521, Egypt

³ Department of Sociology, Faculty of Arts, Beni-Suef University, Beni Suef 62521, Egypt; kayatyashour@art.bsu.edu.eg

* Correspondence: a.salmi@ajman.ac.ae (A.S.); mohamed.abdelgawad3@edu.bsu.edu.eg (M.A.M.)

Abstract

Artificial intelligence (AI) is now woven into university communication, learning, and administration, exposing students, faculty, and professional staff to scalable cybercrime, social engineering, and personal-data risks. This study tested a Compliance Architecture–Legality–Trust–Vigilance (CALT–V) model in which legal literacy and AI-enabled risk perception act as upstream orientations, trust as a calibration mechanism, vigilance as the proximal behavioral-translation layer, and digital safety behaviors as the outcome. A cross-sectional, role-stratified, non-probability survey of 710 UAE higher-education stakeholders (300 students, 200 faculty members, and 210 administrative/professional staff) measured self-reported legal literacy, objective legal knowledge, AI risk perception, institutional and technology/AI trust, vigilance, digital safety behaviors, and perceived governance conditions. Respondents reported moderate objective legal knowledge alongside moderate-to-high self-reported literacy, risk awareness, trust, vigilance, and safety behavior. Confirmatory factor analysis and structural equation modeling were estimated with subscale-level indicators in IBM SPSS AMOS 29.0 (IBM Corp., Armonk, NY, USA), using maximum-likelihood estimation, complete scale-level cases, and 5000-sample bias-corrected bootstrapping for indirect and total effects. Because the model is multidimensional and the sample is role-balanced but non-probability, the results are read as cautious model-based evidence rather than definitive population evidence. AI risk perception was associated with digital safety behaviors both directly and indirectly through trust and vigilance, whereas trust was linked to safety mainly through vigilance. Legal literacy was associated with trust, but its total association with digital safety behavior was not statistically reliable. Rather than claiming that human oversight is a new idea, the findings specify how trust-with-verification can be operationalized in UAE university cyber-governance through concrete verification, reporting, and incident-escalation routines.



Academic Editors: Hai Thanh Luong, Simon Mackenzie and Daniel Silverstone

Received: 17 May 2026

Revised: 18 July 2026

Accepted: 27 July 2026

Published: 26 August 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.

Keywords: legal literacy; AI risk perception; digital safety behaviors; trust calibration; vigilance; cyber hygiene; cybercrime; personal-data protection; UAE universities; higher-education governance

1. Introduction

Artificial intelligence (AI) is now embedded in university communication, learning-management systems, administrative portals, search tools, assessment workflows, and everyday digital decisions. This expansion improves access, responsiveness, and productivity, but it also widens exposure to cyber-enabled manipulation. AI-supported deception

can imitate institutional tone, automate persuasive messaging, personalize phishing scripts, generate synthetic voices or images, and circulate plausible content at scale. Recent reviews of generative-AI-enabled social engineering show that realistic content creation, precise targeting and personalization, and automated attack infrastructure can substantially amplify phishing and related deception risks. In higher education, where users routinely act on course notices, registration requests, financial communications, credential prompts, and research-administration messages, these risks are not peripheral; they are built into the ordinary infrastructure of academic life and therefore call for behavioral, legal, and governance responses rather than technical controls alone (Aggarwal et al. 2025; Akdemir and Lawless 2020; Areghan 2023; Islam et al. 2025; Schmitt and Flechais 2024).

The United Arab Emirates offers a particularly instructive setting for this problem. Its universities operate within a national digital-transformation agenda and under a legal architecture that includes Federal Decree-Law No. 34 of 2021 on countering rumors and cybercrimes, and Federal Decree-Law No. 45 of 2021 concerning the protection of personal data. These instruments set expectations for responsible digital conduct, lawful data handling, data-subject rights, controller and processor duties, and institutional due diligence. Practitioner commentary is used only as a supplementary explanation of the personal-data-protection framework, not as the primary authority for legal propositions. The existence of legal standards, however, does not guarantee that university stakeholders can recognize, interpret, and act on them under fast-moving, AI-mediated risk conditions (United Arab Emirates 2021a, 2021b; see also Bird and Bird 2021).

This gap matters because digital safety rests on situated judgments. Users must decide whether a message is credible, whether a request for data is legitimate, whether a disclosure is proportionate, and whether an unusual interaction warrants verification or reporting. Such judgments are shaped by legal literacy, threat appraisal, trust in institutional and technological systems, and the presence or absence of vigilant checking routines. Protection-motivation and cybersecurity-behavior research shows that perceived severity, vulnerability, response efficacy, and usable response pathways influence protective action, while trust-in-automation research warns that confidence becomes harmful when it hardens into uncritical reliance (Boss et al. 2015; Ifinedo 2012; Lee and See 2004; Stanton et al. 2005).

The present study addresses this problem through the Compliance Architecture–Legality–Trust–Vigilance (CALT–V) model. The model treats legal literacy and AI risk perception as upstream orientations, trust as a calibration mechanism, vigilance as the proximal behavioral-translation layer, and digital safety behaviors as the practical outcome. Rather than assuming that legal awareness or risk concern automatically produces safer conduct, CALT–V asks how awareness, perceived risk, and trust are converted into verification, cautious information handling, secure access habits, and reporting. The study does not set out to establish that humans should stay involved when AI is used, since that principle is already well accepted. Its contribution is more specific: an empirical test, grounded in one legal and institutional context, of how legal literacy, AI risk perception, trust, and vigilance fit together among university stakeholders in the UAE, a setting that remains comparatively underrepresented in the cyber-behavior and higher-education governance literature.

2. Research Questions

RQ1. What are the descriptive levels of legal literacy, AI-enabled cyber-risk perception, trust, vigilance, digital safety behaviors, and perceived governance conditions among UAE university stakeholders?

RQ2. How are legal literacy, AI risk perception, trust, vigilance, digital safety behaviors, and governance conditions associated at the construct level?

RQ3. Does the CALT-V structural model account for direct and indirect pathways linking legal literacy and AI risk perception to digital safety behaviors through trust and vigilance?

RQ4. Do the focal constructs differ by institutional role (students, faculty members, and administrative/professional staff), and does a pooled student-versus-employee comparison alter the substantive interpretation at the broad composite level?

3. Literature Review

3.1. AI-Enabled Cyber Risk and Cyber Hygiene in University Ecosystems

AI-mediated cyber risk is increasingly understood as a socio-behavioral problem rather than a purely technical one. Cyber-enabled offending often succeeds not because systems are unprotected, but because users are persuaded to authorize access, disclose information, or treat falsified content as genuine. Systematic reviews of generative-AI-enabled deception stress that social-engineering risks are amplified by realistic content generation, precise targeting and personalization, automated attack infrastructure, and human-factor weaknesses in verification (Jabir et al. 2025; Schmitt and Flechais 2024). Universities are especially exposed because their routines demand high volumes of email exchange, document sharing, identity confirmation, learning-platform access, and time-sensitive responses. A deceptive message can borrow familiar institutional cues and reach users in a setting where speed, compliance, and habitual trust are already the norm (Akdemir and Lawless 2020; Areghan 2023; Oerlemans et al. 2024).

AI heightens this vulnerability by manufacturing credibility at scale. Large-language-model outputs reproduce professional register, institutional wording, and context-specific urgency, while synthetic media unsettle voice- and image-based trust cues. Users who once relied on fluency, formatting, logos, or familiar style as markers of authenticity now face contexts in which those signals are cheap to imitate. Cyber-hygiene guidance has therefore shifted toward practical habits such as multifactor authentication, link and sender verification, password management, data minimization, incident reporting, and out-of-band confirmation, rather than generic awareness alone (Jabir et al. 2025; National Institute of Standards and Technology 2024; Stanton et al. 2005).

Higher education also presents distinctive risk surfaces. Students receive registration, fee, scholarship, assessment, and learning-platform messages; faculty members handle manuscripts, grades, advising records, research data, and student communications; and administrative or professional staff manage finance, procurement, human resources, credentials, and personal records. These exposures differ in operational detail, yet they share a common behavioral challenge: users must recognize when an apparently legitimate digital request should be slowed down, verified through a second channel, or escalated to an institutional reporting pathway.

3.2. Protection Motivation, AI Risk Perception, and Protective Security Behavior

Risk perception is central to protective conduct because users are more likely to act when a threat feels concrete, plausible, and personally relevant rather than abstract or distant. Protection-Motivation Theory explains protective action through threat appraisal and coping appraisal, encompassing perceived severity, vulnerability, response efficacy, self-efficacy, and response costs (Rogers 1975). Reviews of information-security compliance show that Protection-Motivation Theory, the Theory of Planned Behavior, and General Deterrence Theory are frequently used to explain cyber-information-security compliance and violation behavior in organizations (Sulaiman et al. 2022). Information-security research has extended this logic to policy compliance, security behavior, and fear-appeal designs,

showing that risk awareness is most effective when it is paired with feasible protective actions and clear procedural cues (Boss et al. 2015; Ifinedo 2012).

In AI-mediated environments, perceived risk still involves severity and vulnerability but adds a further element: credibility amplification, the sense that automation, personalization, synthetic realism, and fluent language make deception harder to detect. This matters because AI does not simply increase the number of possible attacks; it changes the signals users rely on when judging whether communication is authentic. Work on AI risk perception and public attitudes toward regulation suggests that how risk is framed can shape concern, governance preferences, and willingness to adopt protective routines (Bullock et al. 2025; Neyazi et al. 2023; Twomey et al. 2023).

3.3. Legal Literacy, Legal Consciousness, and the Translation of Rules into Practice

Legal literacy is commonly described as knowledge of rights, obligations, procedures, and legal reasoning, while legal-consciousness scholarship emphasizes how law is experienced, interpreted, and enacted in everyday life (Ewick and Silbey 1998; Zariski 2014). In digitally mediated universities, legal literacy has to function as applied judgment. Stakeholders need to understand not only that personal data are protected or that cyber misconduct is unlawful, but also how those principles should guide decisions about links, account requests, data disclosure, reporting, and evidence preservation.

The UAE's cybercrime and personal-data-protection frameworks make this applied dimension especially salient. Federal Decree-Law No. 34 of 2021 and Federal Decree-Law No. 45 of 2021 establish the official compliance environment for cybercrime, rumors and online misconduct, personal-data processing, data-subject rights, controller and processor duties, and data-protection governance. Their practical force depends in part on whether users can translate legal standards into routine institutional behavior. A policy that says users should protect data is less actionable than a workflow specifying how to confirm an unexpected credential request, where to report suspicious messages, what information should never be shared through insecure channels, and how to preserve evidence when AI-generated deception is suspected (United Arab Emirates 2021a, 2021b; Bird and Bird 2021).

This study therefore treats legal literacy as a compliance architecture rather than a single awareness score. Rights literacy, obligations literacy, and procedural literacy are modeled as self-reported latent indicators because they capture users' perceived ability to interpret and apply legal-compliance information. Objective legal knowledge is kept as a separate descriptive and correlational indicator rather than part of the latent legal-literacy construct, because it relies on dichotomous accuracy scoring, uses a different measurement format, and reflects factual recognition rather than perceived rights, obligations, and procedural readiness. This separation addresses the possibility that respondents may feel legally literate while still lacking specific legal knowledge.

3.4. Trust in AI, Overreliance, and Calibrated Vigilance

Trust is indispensable in digital institutions, yet it is not inherently protective. Institutional trust is confidence that the university communicates transparently, handles incidents fairly, and provides legitimate reporting pathways. Technology or AI trust is the perceived reliability of systems, recommendations, automated cues, and AI-supported services. Trust in automation can guide appropriate reliance in complex settings, but it can also turn maladaptive when users lean on systems or institutional-looking messages without monitoring them critically (Lee and See 2004).

This is the trust paradox. Trust supports orderly participation in digital systems, but too much of it can suppress checking, encourage compliance on autopilot, and create openings when deceptive content looks authoritative. Research on human–AI trust and

overreliance shows that confidence in AI-mediated outputs can carry accuracy and safety costs when users defer judgment instead of scrutinizing uncertain or unusual outputs (Ferrario et al. 2020; Kaplan et al. 2023; Klingbeil et al. 2024; Wong et al. 2024).

Vigilance addresses this behavioral gap. It combines normative endorsement of verification principles with the enacted habit of pausing, checking, confirming, and reporting. In the CALT-V model, vigilance is not generalized suspicion but trust-with-verification: users should trust institutional reporting channels and official governance processes while still verifying unusual requests, resisting urgency pressure, and escalating uncertain interactions through transparent workflows. This is consistent with risk-management frameworks that prize governable, measurable, and auditable routines over passive trust (Habbal et al. 2024; Laux et al. 2023; National Institute of Standards and Technology 2023, 2024).

3.5. Governance Clarity and Enforcement as Contextual Conditions

Governance clarity and enforcement matter because individual vigilance is easier to enact when users know which rules apply, which channels are official, and what consequences or support follow an incident. In this study, these dimensions are treated as contextual and descriptive governance indicators rather than focal SEM constructs. The reasoning is both theoretical and statistical: governance clarity and enforcement overlap strongly with institutional trust and could obscure the trust-calibration mechanism that the study is designed to test. They therefore help contextualize RQ1 and RQ2 and inform practical implications, while the structural test stays focused on legal literacy, AI risk perception, trust, vigilance, and digital safety behavior.

3.6. Positioning the Present Study

Taken together, the literature points to an integrated explanatory pathway. Legal literacy can make digital obligations and protections intelligible; AI risk perception can supply motivational urgency; trust can either anchor institutional legitimacy or drift into overreliance; governance clarity can make response pathways visible; and vigilance can turn these orientations into action. What remains underexamined is how these constructs operate together among university stakeholders within a specific legal and governance environment. The present study addresses that gap by testing the CALT-V model in UAE higher education and by combining descriptive profiling, role-based comparisons, construct-level correlations, CFA, and SEM with an explicit account of methodological limits. Its value lies in extending a largely international cyber-behavior discussion into a UAE university context, not in presenting the human-in-the-loop principle as a new cybersecurity idea.

4. Theoretical Framework: The CALT-V Model

The CALT-V model organizes the study around five focal constructs: legal literacy, AI risk perception, trust, vigilance, and digital safety behaviors. Legal literacy represents the compliance architecture through which respondents interpret rights, obligations, and procedures. AI risk perception captures threat appraisal tied to severity, vulnerability, and AI-enabled credibility amplification. Trust represents confidence in institutional governance and in technology- or AI-mediated processes, and vigilance captures verification norms and protective monitoring routines. Digital safety behaviors form the behavioral endpoint, including secure authentication, cautious information handling, and reporting-related action.

Figure 1 sets out the broader conceptual logic of CALT-V. It depicts trust as a calibration mechanism whose protective value depends on whether it is expressed as trust-with-verification rather than trust-without-checking. In the empirical test, this logic is evaluated

through a mediation-centered structural model: legal literacy and AI risk perception are specified as upstream correlates of trust; trust is modeled as a predictor of vigilance; vigilance is modeled as a proximal predictor of digital safety behaviors; and AI risk perception is allowed a direct path to digital safety because risk perception may motivate protective action on its own. The figure provides the theoretical architecture, while the SEM tests the central direct and indirect pathways that the cross-sectional data can support. Latent moderation and role-invariance tests are left to future research.

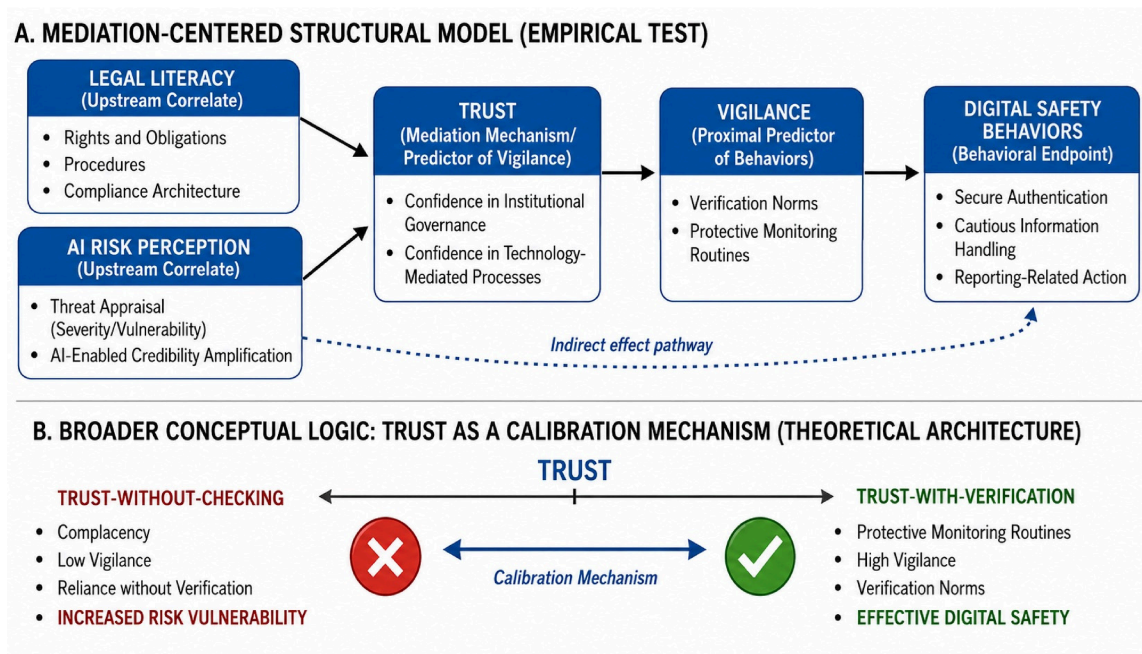


Figure 1. The Compliance Architecture–Legality–Trust–Vigilance (CALT–V) conceptual model. AI = artificial intelligence.

Hypotheses

The four research questions are addressed through complementary analyses. RQ1 and RQ2 are descriptive and correlational: they establish the levels of the focal constructs and their construct-level associations, and they are examined through descriptive statistics and correlations rather than through formal hypotheses. RQ3 and RQ4 are inferential: RQ3 concerns the structural pathways linking legal literacy and AI risk perception to digital safety behaviors, and RQ4 concerns role-based differences. The hypotheses below operationalize RQ3 by specifying the trust-calibrated vigilance mechanism, whereas RQ4 is examined through role comparisons rather than a directional hypothesis. The hypotheses are grouped into four theoretically grounded propositions that reflect the core logic of that mechanism, and individual path estimates are reported in the Results to preserve analytic transparency.

H1. Upstream orientations hypothesis: Legal literacy and AI risk perception are positively associated with trust, because legally and risk-informed stakeholders are expected to find institutional and technological governance more legible.

H2. Trust-calibration hypothesis: Trust is positively associated with vigilance, because protective trust should be expressed as verification-oriented monitoring rather than passive reliance.

H3. Behavioral-translation hypothesis: Vigilance is positively associated with digital safety behaviors and is expected to be a stronger proximal predictor than trust alone; AI risk perception is also expected to retain a direct positive association with digital safety behaviors.

H4. *Indirect pathway hypothesis: Trust has a positive indirect association with digital safety behaviors through vigilance; AI risk perception and legal literacy may also show indirect associations with digital safety behaviors through the trust–vigilance pathway, although the indirect pathway is expected to be stronger for AI risk perception than for legal literacy.*

5. Materials and Methods

5.1. Research Design and Setting

The study used a cross-sectional, explanatory-correlational design to examine the CALT–V model in UAE university environments. This design suited a study focused on associations among latent constructs and on modeled direct and indirect pathways linking legal literacy and AI risk perception to digital safety behaviors. The university context is analytically important because institutional email, learning-management systems, student portals, cloud services, research-administration systems, and administrative workflows generate frequent data-handling and verification decisions under conditions of digital urgency.

The design does not support causal inference. SEM paths are therefore interpreted as theoretically specified associations rather than evidence that one construct caused another. The study was designed to provide a model-based explanatory test and a descriptive profile of UAE higher-education stakeholders, not a nationally representative prevalence estimate.

The sample is substantial for a university survey, but it should be read against the breadth of the model. CALT–V includes several closely related constructs and multiple subdimensions, and for that reason, the CFA and SEM were estimated using theoretically defined subscale-level indicators rather than a larger item-level model. This choice improved parsimony and reduced parameter burden, but it also means the analysis is best treated as a cautious model-based test that requires replication in larger and more institutionally diverse samples.

5.2. Participants, Sampling, and Eligibility

The analytic sample comprised 710 respondents from UAE higher-education settings: 300 students, 200 faculty members, and 210 administrative or professional staff. These groups were chosen because they engage with university digital systems in different ways. Students navigate registration, learning, assessment, financial, and credential communications; faculty members manage teaching, advising, assessment, research, and student records; and administrative or professional staff coordinate institutional procedures and data-intensive services.

Recruitment followed a role-stratified non-probability structure. The aim was to secure adequate representation from the three stakeholder categories while relying on convenience recruitment through institutional digital channels and professional and student networks. Eligibility required that respondents be currently enrolled or employed in a UAE university context, use at least one official digital channel connected to university operations, and be able to complete the English-language questionnaire. The questionnaire was administered in English because English is widely used in UAE higher-education administration and instruction; no Arabic translation or back-translation procedure was therefore applied.

Because the survey link was circulated through decentralized online channels and no identifiable invitation logs were retained, a precise response rate could not be calculated. The sample should therefore be read as role-balanced but non-representative. Participants were recruited from more than one UAE higher-education institution; however, the number of participating universities and their institutional type (for example, public or private) are not disclosed, because the ethics approval and the anonymous recruitment procedure were designed to preserve institutional anonymity and no institution-identifying

information was collected that would support a reliable classification. Institutional and emirate-level coverage is therefore described only in broad terms, and the findings should not be generalized to all UAE universities without caution.

5.3. Measures and Scale Development

The survey instrument operationalized the CALT-V constructs through multi-item scales. Unless otherwise noted, attitudinal items used a five-point agreement scale and behavior-oriented items used a recent-period frequency format. The instrument combined constructs adapted from prior work on cybersecurity behavior, protection motivation, legal literacy and legal consciousness, and trust in automation and AI with study-specific items tailored to UAE university cyber-governance and AI-mediated deception (Boss et al. 2015; Ewick and Silbey 1998; Ifinedo 2012; Lee and See 2004; Rogers 1975; Stanton et al. 2005; Zariski 2014).

Legal literacy was measured through three self-reported subscales: rights literacy, obligations literacy, and procedural literacy. A separate six-item objective legal-knowledge check captured factual recognition of UAE cybercrime and personal-data-protection principles. This objective component was excluded from the latent legal-literacy construct because it used dichotomous scoring, assessed factual accuracy rather than perceived procedural readiness, and produced a narrower 0–6 range. It was instead reported descriptively and entered into the correlation matrix as LLK. The distinction is important because legal literacy may involve both perceived readiness to act and objectively correct legal knowledge, and the two need not coincide.

AI risk perception measured perceived severity, perceived vulnerability, and AI-enabled credibility amplification. Trust measured institutional trust, technology/AI trust, and an overreliance indicator. For descriptive trust scoring, overreliance items were reverse-scored before the overall trust total was computed, so that higher total trust reflected more calibrated and less overreliant trust. For transparency in the CFA and SEM, the overreliance subscale was retained under its original substantive label as a reverse-oriented indicator, which explains the negative standardized loading shown in the measurement and structural figures. Vigilance measured normative and behavioral vigilance. Digital safety behaviors measured adoption and avoidance actions, information hygiene, and verification and reporting routines. Governance conditions were represented by governance clarity and enforcement. Key item domains are summarized in Appendix A, and the full instrument should be supplied as Supplementary Material for reproducibility.

Because several items were adapted to the UAE legal and university context and others were newly developed for this research, the scales are best regarded as study-specific measures with internal validation evidence. The present manuscript reports construct-level descriptive statistics, CFA loadings, composite reliability, AVE, and discriminant-validity interpretation. Cronbach's alpha or McDonald's omega should be reported in the Supplementary Instrument File if item-level response matrices are supplied, because the CFA and SEM here used subscale indicators rather than individual items.

5.4. Data Collection and Ethics

Data were collected through an anonymous online survey. Participation was voluntary, and respondents received an informed consent statement describing the study's purpose, confidentiality protections, and the right to withdraw before submission. The survey did not request passwords, account identifiers, IP logs, names, student numbers, employee numbers, or other credentials. Because the topic concerned cybercrime and personal data, the questionnaire focused on perceptions, knowledge, and protective routines rather than soliciting admissions of wrongdoing.

Response-quality screening applied minimum completion thresholds and attention to careless-response patterns before the final analytic sample was fixed at $N = 710$. Missing or incomplete cases were excluded before scale totals were computed, so the final dataset contains complete scale-level observations for all focal constructs.

5.5. Analytic Strategy

The analytic sequence moved through five linked stages. Descriptive statistics first summarized the central tendency and dispersion of the focal constructs. Role-based mean differences were then examined using one-way analysis of variance (ANOVA) across students, faculty members, and administrative/professional staff, with a pooled student-versus-employee Welch comparison retained as a sensitivity analysis because earlier analytic framing contrasted learners with university employees. Pearson product-moment correlations next assessed construct-level associations, including governance clarity and enforcement as contextual variables.

A confirmatory factor analysis then evaluated the measurement structure for legal literacy, AI risk perception, trust, vigilance, and digital safety behaviors, using subscale-level indicators rather than individual item indicators. This reduced model complexity and aligned the SEM with the theoretical subdimensions, but it also means that very strong fit indices should be read as evidence of coherent composite measurement rather than proof of flawless item-level validity. Composite reliability (CR) and average variance extracted (AVE) were computed from the standardized CFA loadings, and discriminant validity was interpreted using the Fornell–Larcker logic, comparing the square root of AVE with interconstruct correlations (Fornell and Larcker 1981).

Finally, SEM tested the CALT–V pathways using maximum-likelihood estimation in IBM SPSS AMOS 29.0 (IBM Corp., Armonk, NY, USA), while descriptive statistics, correlations, ANOVAs, and Welch tests were conducted in IBM SPSS Statistics 29.0. The CFA and SEM used subscale-level indicators rather than item-level indicators or single scale composites, and missing or incomplete responses were handled by complete-case exclusion before scale totals were computed, so the analyzed dataset contained complete scale-level observations for all focal constructs. Before estimation, the subscale indicators were screened for distributional problems; univariate skewness and kurtosis fell within conventionally accepted limits for maximum-likelihood SEM, and no indicator showed extreme departures from normality. Because maximum-likelihood estimation nonetheless assumes multivariate normality, an assumption rarely met exactly in behavioral data, direct effects were interpreted with standardized coefficients, and indirect and total effects were estimated with bias-corrected bootstrapping (5000 resamples) and 95% confidence intervals. Bootstrapping does not rely on the normality assumption and therefore provides a robust check on the mediation estimates that are central to the model. Given the cross-sectional design, structural pathways are interpreted as modeled associations rather than causal effects. No custom computer code was developed for this study.

Common method bias was addressed procedurally through anonymity, separation of construct sections, mixed item domains, objective legal-knowledge items, and frequency-based behavior items (Podsakoff et al. 2003). The study did not, however, include a marker variable or report a full item-level common-method-factor test. This residual risk is acknowledged in the Limitations section rather than treated as resolved.

5.6. Model Respecification

The initial SEM represented a theoretically generous version of CALT–V that allowed both mediated and direct associations to be examined. Respecification was limited and theory guided. Compared with the initial model, the final retained SEM clarified the

trust-calibrated pathway: it kept the core trust-to-vigilance and vigilance-to-digital-safety paths, retained the direct AI-risk-perception-to-digital-safety path justified by Protection-Motivation Theory, and preserved the theoretically important trust-to-digital-safety path as a non-significant direct test rather than deleting it after estimation. No residual covariances were added for statistical improvement alone, and no role- or governance-related paths were added post hoc. Because the two models were parameterized with the same number of free parameters, their degrees of freedom are identical; the comparison should therefore be read as a comparison of two theoretically distinct specifications of equal complexity, not as evidence of added model flexibility. The final SEM is a more transparent representation of the same theoretical architecture, not an exploratory search for the best-fitting model.

6. Results

Addressing RQ1, Table 1 shows that participants reported generally moderate-to-high levels across the study constructs. Objective legal-literacy knowledge was moderate, with respondents answering an average of 3.86 of 6 items correctly ($SD = 1.19$), whereas self-reported legal literacy was relatively stronger ($M = 59.65$, $SD = 11.19$). This gap supports distinguishing factual legal knowledge from perceived rights, obligations, and procedural literacy. AI risk perception was elevated ($M = 55.63$, $SD = 10.85$), with closely aligned severity, vulnerability, and AI-amplification subscales. Trust ($M = 62.61$, $SD = 6.42$), vigilance ($M = 55.88$, $SD = 10.70$), and digital safety behaviors ($M = 63.00$, $SD = 12.17$) all fell within moderate-to-high ranges, while governance conditions were viewed as moderately favorable ($M = 38.13$, $SD = 6.47$). Overall, respondents appeared positively oriented towards cyber-governance and safety, though their objective legal knowledge lagged behind their self-perceived literacy and protective dispositions.

Table 1. Descriptive statistics for the focal study variables ($N = 710$).

Measure	Possible Range	N	M	SD	Median
Legal Literacy knowledge (LL-K total)	0–6	710	3.86	1.19	4.00
Rights literacy (LL-R)	5–25	710	17.47	3.62	18.00
Obligations literacy (LL-O)	5–25	710	17.41	3.54	18.00
Procedural literacy (LL-P)	6–30	710	20.91	4.29	21.00
Overall legal literacy (LL total)	16–86	710	59.65	11.19	60.00
Risk severity (RP-S)	5–25	710	17.39	3.63	17.00
Risk vulnerability (RP-V)	5–25	710	17.31	3.63	17.00
AI amplifiers (RP-AI)	6–30	710	20.94	4.19	21.00
Overall AI risk perception (RP total)	16–80	710	55.63	10.85	56.00
Institutional trust (T_inst)	7–35	710	24.22	4.98	24.00
Technology/AI trust (T_AI)	6–30	710	20.77	4.28	21.00
Overreliance (T_over)	5–25	710	17.63	3.17	18.00
Overall trust (T total)	18–90	710	62.61	6.42	63.00
Normative vigilance (V_norm)	8–40	710	27.91	5.51	28.00
Behavioral vigilance (V_beh)	8–40	710	27.97	5.50	28.00
Overall vigilance (V total)	16–80	710	55.88	10.70	56.00
Adoption/avoidance actions (DSB-AA)	5–25	710	17.51	3.62	18.00
Information hygiene (DSB-IH)	6–30	710	21.02	4.22	21.00
Verification & reporting (DSB-VRR)	7–35	710	24.47	4.92	25.00
Overall digital safety behaviors (DSB total)	18–90	710	63.00	12.17	63.00
Governance clarity (GC total)	6–30	710	20.83	4.01	21.00
Enforcement (ENF total)	5–25	710	17.30	3.13	18.00
Overall governance conditions	11–55	710	38.13	6.47	38.00

Note. LL = legal literacy; RP = AI risk perception; T = trust; V = vigilance; DSB = digital safety behaviors. LL-K is the sum of six dichotomous items; other scale totals sum Likert-type items, with higher scores indicating higher levels of the construct.

6.1. Role-Based Comparisons

RQ4 asked whether the focal constructs differed by institutional role. The primary omnibus comparison used one-way ANOVA across students, faculty members, and administrative/professional staff. As Table 2 shows, no statistically significant role-based differences emerged for legal-literacy knowledge, legal-literacy total, AI risk perception, trust, vigilance, digital safety behaviors, governance clarity, enforcement, or overall governance conditions. Effect sizes were uniformly trivial (η^2 values at or below 0.005), indicating that role membership explained negligible variance in these broad composite measures.

Table 2. Omnibus role-based one-way ANOVAs for the focal composite constructs.

Measure	F(2, 707)	p	η^2
Legal literacy knowledge	0.081	0.922	<0.001
Legal literacy (total)	0.112	0.895	<0.001
AI risk perception (total)	0.063	0.939	<0.001
Trust (total)	1.360	0.257	0.004
Vigilance (total)	0.203	0.816	<0.001
Digital safety behaviors (total)	1.602	0.202	0.005
Governance clarity	1.085	0.339	0.003
Enforcement	0.056	0.946	<0.001
Governance conditions (total)	0.341	0.711	<0.001

Note. Role groups were students ($n = 300$), faculty members ($n = 200$), and administrative/professional staff ($n = 210$). η^2 values indicate negligible role-based variance across the reported composites.

A pooled student-versus-employee sensitivity analysis produced the same conclusion. Table 3 reports Welch tests comparing students with faculty and staff combined. No composite construct differed significantly between the two groups. The largest observed contrast was in digital safety behaviors, for which employees reported a modestly higher mean than students, but the difference remained non-significant and substantively small, $t(644.3) = -1.28$, $p = 0.200$, $g = -0.10$. These converging results justify retaining pooled full-sample structural analyses. They should not, however, be read as evidence of full measurement invariance across roles, because invariance testing was beyond the scope of the present dataset and model specification.

Table 3. Pooled student-versus-employee sensitivity analyses using Welch tests.

Measure	Students ($n = 300$) M $\pm$ SD	Faculty/Staff ($n = 410$) M $\pm$ SD	Δ (Students – Employees)	Welch t(df)	p	Hedges g
Legal literacy knowledge	3.87 $\pm$ 1.16	3.85 $\pm$ 1.22	+0.03	0.30 (662.2)	0.764	+0.02
Legal literacy (total)	59.42 $\pm$ 10.76	59.81 $\pm$ 11.50	−0.39	−0.46 (666.9)	0.643	−0.03
AI risk perception (total)	55.77 $\pm$ 11.39	55.53 $\pm$ 10.45	+0.24	0.29 (611.4)	0.773	+0.02
Trust (total)	62.88 $\pm$ 6.46	62.41 $\pm$ 6.39	+0.47	0.96 (640.5)	0.338	+0.07
Vigilance (total)	56.09 $\pm$ 10.89	55.73 $\pm$ 10.58	+0.36	0.44 (633.6)	0.657	+0.03
Digital safety behaviors (total)	62.32 $\pm$ 12.16	63.50 $\pm$ 12.16	−1.19	−1.28 (644.3)	0.200	−0.10
Governance conditions (total)	38.22 $\pm$ 6.41	38.06 $\pm$ 6.52	+0.16	0.32 (650.6)	0.746	+0.02

Note. Δ is calculated as Students – Employees. Employees combine faculty members and administrative/professional staff. Hedges g follows the same direction as Δ . All tests are two-tailed.

6.2. Construct-Level Correlations

Addressing RQ2, Table 4 presents the construct-level correlation matrix. Legal literacy total correlated positively with AI risk perception ($r = 0.290$), trust ($r = 0.277$), vigilance ($r = 0.385$), and digital safety behaviors ($r = 0.379$), all $p < 0.01$. AI risk perception was positively associated with trust ($r = 0.247$), vigilance ($r = 0.278$), and digital safety behaviors ($r = 0.293$). Trust correlated more strongly with vigilance ($r = 0.376$) than with digital

safety behaviors ($r = 0.235$), suggesting that trust may operate more proximally through a vigilance mechanism than as a direct behavioral driver.

Table 4. Construct-level correlation matrix.

	LLK	LL	RP	T	V	DSB	GC	ENF
LLK	1							
LL	0.332 **	1						
RP	0.108 **	0.290 **	1					
T	0.079 *	0.277 **	0.247 **	1				
V	0.126 **	0.385 **	0.278 **	0.376 **	1			
DSB	0.143 **	0.379 **	0.293 **	0.235 **	0.453 **	1		
GC	0.057	0.264 **	0.206 **	0.752 **	0.291 **	0.224 **	1	
ENF	0.052	0.225 **	0.183 **	0.700 **	0.288 **	0.217 **	0.636 **	1

Note. $N = 710$. Values are Pearson product–moment correlations. LLK = legal literacy knowledge; LL = legal literacy total; RP = AI risk perception total; T = trust total; V = vigilance total; DSB = digital safety behaviors total; GC = governance clarity; ENF = enforcement. * $p < 0.05$. ** $p < 0.01$.

Vigilance showed the largest bivariate association with digital safety behaviors ($r = 0.453$, $p < 0.01$), supporting the model’s expectation that verification-oriented self-regulation is the closest correlate of protective conduct. Governance clarity and enforcement were positively associated with legal literacy, vigilance, and digital safety behaviors, but they were excluded from the main SEM because they were conceptualized as contextual governance conditions and because both correlated highly with trust ($r = 0.752$ and $r = 0.700$, respectively). Their role is therefore descriptive and interpretive rather than structural in the final model.

6.3. Measurement Model

The confirmatory factor analysis supported the subscale-level measurement structure used for the structural model. Figure 2 displays the five-factor measurement model, and Table 5 reports its fit indices. Fit was excellent: $\chi^2(67) = 62.91$, $p = 0.619$, $\chi^2/df = 0.939$, goodness-of-fit index (GFI) = 0.988, normed fit index (NFI) = 0.994, comparative fit index (CFI) = 1.000, Tucker–Lewis index (TLI) = 1.001, and root mean square error of approximation (RMSEA) = 0.000. Standardized loadings reported in the figure were strong for the focal subscale indicators. The objective legal-knowledge score (LL-K) was not part of this latent factor model; it was retained as a separate objective measure for descriptive and correlational analyses. The negative loading for the overreliance indicator was consistent with its reverse orientation within the trust construct.

These unusually strong fit indices call for cautious interpretation. CFI values of 1.000, TLI values slightly above 1.000, and RMSEA values of 0.000 can arise when a well-specified model is estimated with a limited number of reliable subscale-level indicators rather than a large set of heterogeneous item-level indicators. The results indicate that the five composite constructs were internally coherent at the subscale level; they should not be read as proof that the item-level measurement model is perfect or that further validation is unnecessary.

Additional reliability and validity evidence was computed from the standardized loadings and is summarized in Table 5a. Composite reliability and AVE were acceptable as strong for legal literacy, AI risk perception, trust, vigilance, and digital safety behaviors. The square roots of AVE for all focal constructs exceeded the largest observed interconstruct correlations among the focal CFA constructs, supporting discriminant validity under the Fornell–Larcker criterion. Governance clarity and enforcement were excluded from the CFA and SEM because they were treated as contextual governance indicators and correlated strongly with trust in the construct-level matrix. These estimates rest on the subscale

indicators and should be supplemented by item-level alpha or omega coefficients when the full instrument is submitted as Supplementary Material.

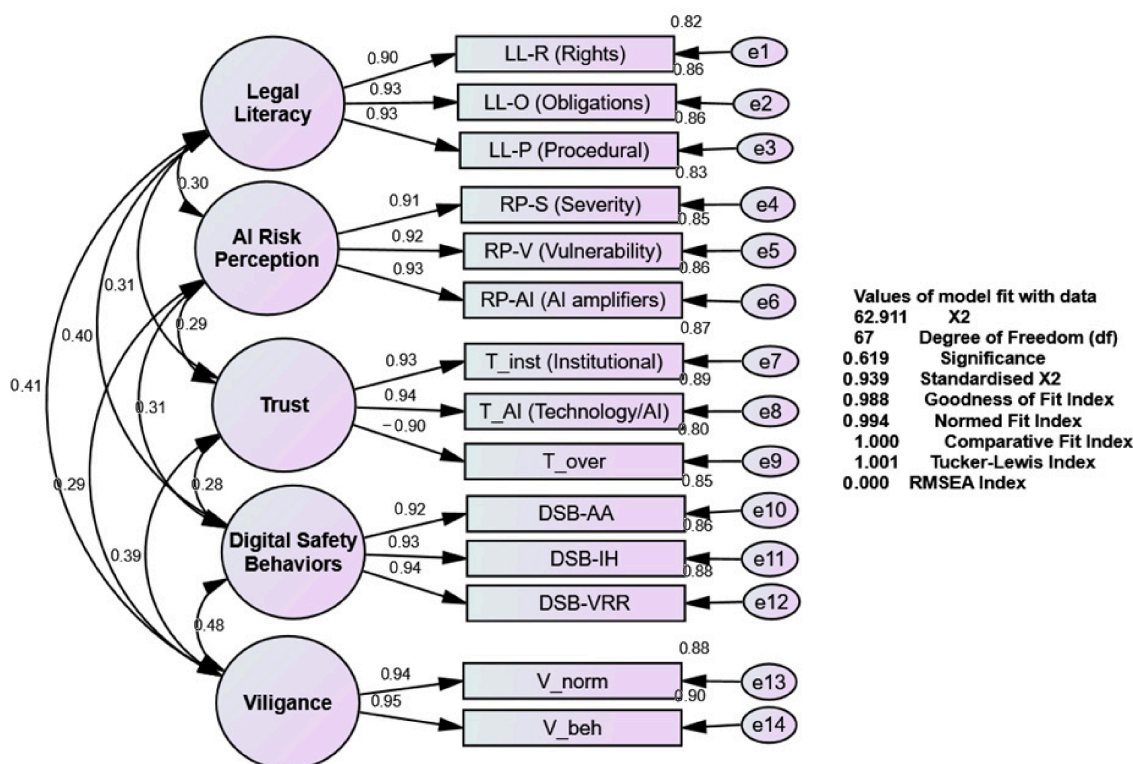


Figure 2. Confirmatory factor analysis measurement model for legal literacy, AI risk perception, trust, digital safety behaviors, and vigilance.

Table 5. (a) Reliability and discriminant-validity evidence for the subscale-level measurement model. (b) Measurement-model fit indices.

(a)					
Construct	Indicators	CR	AVE	$\sqrt{\text{AVE}}$	Largest Focal r
Legal literacy	LL-R, LL-O, LL-P	0.933	0.822	0.907	0.385
AI risk perception	RP-S, RP-V, RP-AI	0.938	0.834	0.913	0.293
Trust	T_inst, T_AI, T_over	0.845	0.660	0.812	0.376
Vigilance	V_norm, V_beh	0.831	0.712	0.844	0.453
Digital safety behaviors	DSB-AA, DSB-IH, DSB-VRR	0.935	0.828	0.910	0.453

(b)		
Fit Statistic	Abbreviation	Value
Chi-square	χ^2	62.911
Degrees of freedom	df	67
Significance (p-value)	p	0.619
Standardized chi-square	χ^2/df	0.939
Goodness-of-Fit Index	GFI	0.988
Normed Fit Index	NFI	0.994
Comparative Fit Index	CFI	1.000
Tucker–Lewis Index	TLI	1.001
Root Mean Square Error of Approximation	RMSEA	0.000

Note. CR = composite reliability; AVE = average variance extracted. Largest focal r is the largest observed correlation with another construct included in the focal CFA/SEM. Governance clarity and enforcement were treated as contextual variables and were not included in this measurement model. Note. Fit indices correspond to the CFA measurement model displayed in Figure 2.

6.4. Structural Model and Hypothesis Tests

The structural model addresses RQ3. The initial SEM provided acceptable-to-good fit, whereas the final retained model showed stronger overall fit and was used for interpretation. Figure 3 presents the initial model, Figure 4 the final model, and Table 6 compares their fit indices. Relative to the initial model, the final SEM showed lower χ^2 , lower χ^2/df , lower RMSEA, higher CFI and TLI, and lower AIC. GFI and NFI were marginally lower in the final specification, but the overall pattern favored the final model as the more efficient representation of the data.

The respecification was conservative and guided by CALT-V theory, the protection-motivation rationale for a direct AI-risk-perception-to-safety path, and the concern that the model should not appear data-driven. The final SEM retained the core trust-to-vigilance and vigilance-to-safety pathway, retained the direct AI-risk-perception-to-safety pathway, and kept the non-significant trust-to-safety direct path for theoretical transparency. It did not add correlated residuals, role paths, or governance paths after fit was inspected. The identical degrees of freedom in Table 6 reflect equal overall model complexity, not an unreported increase in free parameters. The final model preserves the central claim: trust matters for safety mainly when it is translated into vigilance.

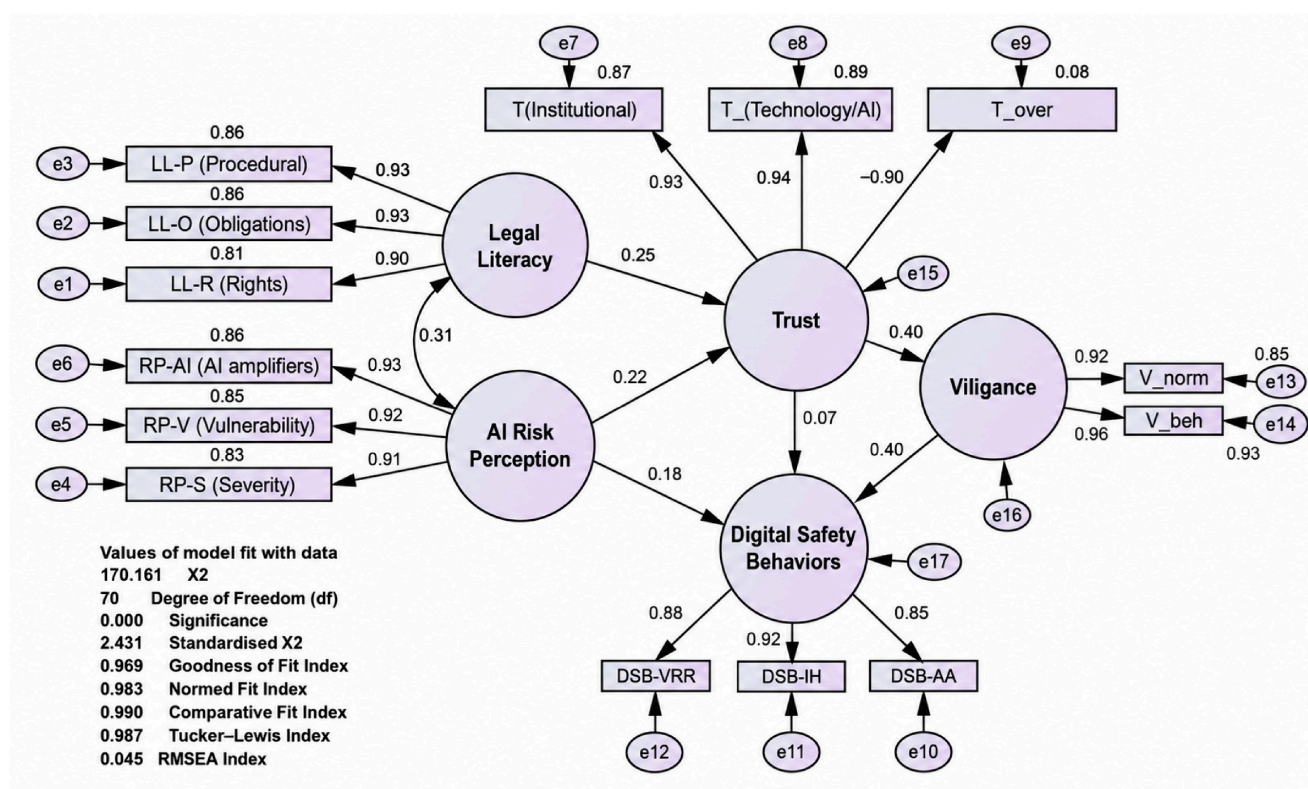


Figure 3. Initial structural equation model linking legal literacy and AI risk perception to trust, vigilance, and digital safety behaviors.

The direct effects from the final SEM are reported in Table 7. The paths from legal literacy to trust ($\beta = 0.132$, $p = 0.023$) and from AI risk perception to trust ($\beta = 0.208$, $p < 0.001$) were significant, supporting the upstream orientations hypothesis (H1). The path from trust to vigilance was significant ($\beta = 0.317$, $p < 0.001$), supporting the trust-calibration hypothesis (H2). The path from vigilance to digital safety behaviors was significant ($\beta = 0.264$, $p < 0.001$), and AI risk perception also retained a direct positive association with digital safety behaviors ($\beta = 0.253$, $p < 0.001$), supporting the behavioral-translation hypothesis (H3). Trust did not have a statistically reliable direct association

with digital safety behaviors ($\beta = 0.048, p = 0.418$), which is theoretically important because it indicates that trust alone is not a sufficient behavioral engine.

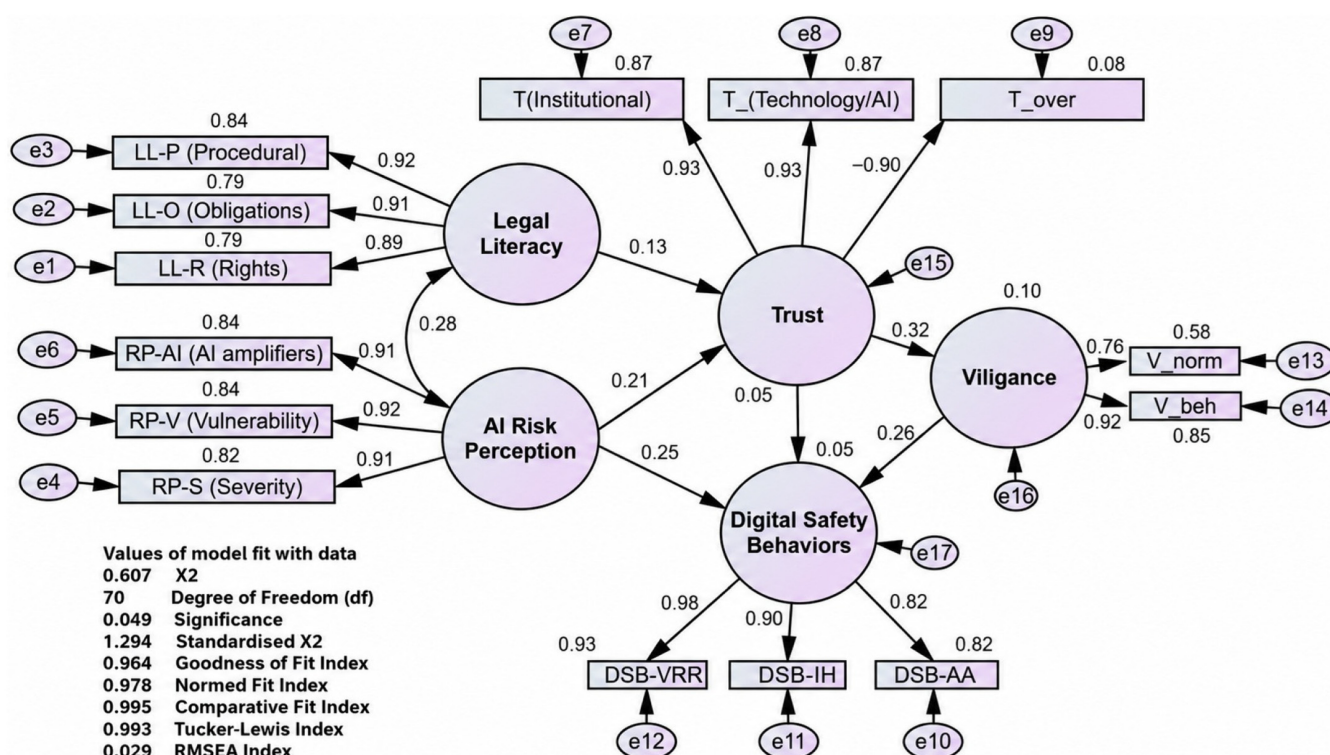


Figure 4. Final retained structural equation model interpreted in the study.

Table 6. Fit indices for the initial and final structural equation models.

Fit Index	Initial SEM	Final SEM
χ^2	170.161	90.607
df	70	70
p	<0.001	0.049
χ^2/df	2.431	1.294
GFI	0.969	0.964
NFI	0.983	0.978
CFI	0.990	0.995
TLI	0.987	0.993
RMSEA	0.045	0.029
AIC	171.434	160.607

Note. The final SEM was retained because it showed stronger approximation and incremental fit on key indices, lower χ^2/df , and lower AIC, while preserving theoretically interpretable pathways. The two models have identical degrees of freedom because the final specification did not increase the number of free parameters. The respecification was theory-guided and did not rely on post hoc correlated error terms.

Table 7. Direct structural effects from the final SEM.

Path	β	B	SE	CR	p
Digital Safety Behaviors $\leftarrow$ AI Risk Perception	0.253	0.225	0.050	4.522	<0.001
Digital Safety Behaviors $\leftarrow$ Trust	0.048	0.032	0.040	0.810	0.418
Digital Safety Behaviors $\leftarrow$ Vigilance	0.264	0.318	0.073	4.343	<0.001
Trust $\leftarrow$ AI Risk Perception	0.208	0.274	0.077	3.568	<0.001
Trust $\leftarrow$ Legal Literacy	0.132	0.190	0.084	2.271	0.023
Vigilance $\leftarrow$ Trust	0.317	0.177	0.037	4.793	<0.001

Note. β = standardized coefficient; B = unstandardized estimate; SE = standard error; CR = critical ratio.

The indirect effects clarify the model's central mechanism. As Table 8 shows, trust had a significant indirect association with digital safety behaviors through vigilance ($\beta_{\text{ind}} = 0.084, p < 0.001$), supporting the indirect pathway hypothesis (H4). AI risk perception also had a smaller but significant indirect association with digital safety behaviors through the modeled trust–vigilance pathway ($\beta_{\text{ind}} = 0.027, p = 0.038$), whereas the analogous indirect association for legal literacy was not statistically reliable ($\beta_{\text{ind}} = 0.017, p = 0.191$). Table 9 further shows that the total association of AI risk perception with digital safety behaviors was substantial and significant ($\beta_{\text{total}} = 0.281, p < 0.001$), while the total association for legal literacy remained non-significant. Legal literacy therefore appears to strengthen institutional and technology trust, whereas AI risk perception and vigilance are more directly behavior-linked in the final model.

Table 8. Standardized indirect effects from the final SEM.

Predictor	Outcome	β_{ind}	Boot SE	95% CI LL	95% CI UL	p (Approx.)
Legal Literacy	Digital Safety Behaviors	0.017	0.013	−0.008	0.042	0.191
AI Risk Perception	Digital Safety Behaviors	0.027	0.013	0.002	0.052	0.038
Trust	Digital Safety Behaviors	0.084	0.023	0.039	0.129	<0.001
Legal Literacy	Vigilance	0.042	0.022	−0.001	0.085	0.056
AI Risk Perception	Vigilance	0.066	0.022	0.023	0.109	0.003

Note. Indirect effects are reported with bootstrapped standard errors and 95% bias-corrected confidence intervals based on 5000 bootstrap resamples as provided in the model output.

Table 9. Standardized total effects on digital safety behaviors in the final SEM.

Predictor	Outcome	β_{total}	Boot SE	95% CI LL	95% CI UL	p (Approx.)
Legal Literacy	Digital Safety Behaviors	0.017	0.013	−0.008	0.042	0.191
AI Risk Perception	Digital Safety Behaviors	0.281	0.054	0.175	0.387	<0.001
Trust	Digital Safety Behaviors	0.131	0.056	0.021	0.241	0.019
Vigilance	Digital Safety Behaviors	0.264	0.060	0.146	0.382	<0.001

Note. Total effects combine direct and indirect contributions where applicable.

7. Discussion

This study tested a trust-calibrated vigilance account of digital safety behaviors among UAE university stakeholders, and its findings speak directly to the four research questions. The descriptive profile for RQ1 showed moderate objective legal knowledge alongside generally favorable self-reported literacy, risk awareness, trust, vigilance, and safety behavior, and the construct-level correlations for RQ2 showed that vigilance was the closest correlate of protective conduct. Addressing RQ3, the structural model supported the central logic of CALT–V in an associational sense: legal literacy and AI risk perception are meaningful upstream orientations, but vigilance is the most proximal behavioral mechanism linking orientation to action, and not all upstream orientations were linked to protective behavior with equal force. The role comparisons for RQ4 revealed no meaningful differences across stakeholder groups. This reading is cautious because the model is multidimensional and the sample, though sizeable, is not a population sample of UAE universities.

7.1. Interpreting the CALT–V Pathway

AI risk perception emerged as the most behaviorally consequential distal construct. It was associated with digital safety behaviors both directly and indirectly through trust and vigilance. This fits protection-motivation reasoning: threat perception is most likely to matter when users regard the threat as concrete, personally relevant, and paired with feasible protective action. In the AI context, the relevant threat is not only conventional

cybercrime severity or vulnerability but also credibility amplification, in which fluent text, institutional tone, synthetic media, personalization, and automation make deception harder to detect (Boss et al. 2015; Bullock et al. 2025; Ifinedo 2012; Jabir et al. 2025; Neyazi et al. 2023; Schmitt and Flechais 2024).

7.2. Trust as a Calibrated Resource Rather than a Direct Behavioral Engine

The most important theoretical contribution concerns trust. Trust was not directly associated with digital safety behaviors in the final SEM, but it was strongly associated with vigilance and contributed to safety behavior indirectly through vigilance. This helps resolve the trust paradox: trust is useful when it supports monitoring, checking, and procedural engagement, and unhelpful when it substitutes for them. The data thus favor a trust-with-verification interpretation over a trust-as-reassurance one.

This aligns with research on trust in automation and human–AI systems. Trust can support appropriate reliance when users understand the system context, but overreliance can undermine accuracy and safety when users defer judgment to apparently authoritative outputs or institutional-looking messages (Kaplan et al. 2023; Klingbeil et al. 2024; Lee and See 2004). In university cyber-governance, trust should therefore be designed as a calibrated institutional resource: stakeholders should be able to trust reporting channels, communication protocols, and governance procedures while still being encouraged to verify unusual requests and resist persuasive urgency.

7.3. Legal Literacy: Necessary but Not Sufficient

Legal literacy was positively associated with trust, suggesting that understanding rights, obligations, and procedures can make governance systems more legible and institutionally credible. Yet legal literacy showed no statistically reliable total association with digital safety behaviors in the final model. This does not diminish the importance of legal literacy; it clarifies its function. Legal understanding may build interpretive readiness, but it does not automatically produce behavioral compliance under time pressure, social-engineering pressure, or persuasive AI-mediated cues.

The descriptive gap between objective legal knowledge and self-reported legal literacy sharpens this point. Respondents' objective legal-knowledge scores were moderate while their self-reported literacy scores were higher. This suggests that universities should not rely on perceived confidence or policy familiarity alone. Legal literacy programming should include short factual checks, scenario-based application, and clear procedural mapping, so that stakeholders know what the law expects and what institutional actions should follow.

7.4. Governance Conditions and Institutional Design

Governance clarity and enforcement were positively associated with the focal constructs, but their strongest correlations were with trust, which supports treating them as contextual conditions rather than central SEM pathways in the present model. Conceptually, governance clarity is the institutional infrastructure that makes vigilance possible: users are more likely to verify and report when they know which channels are official, which requests are unusual, where to report suspicious messages, and what support or enforcement follows.

The practical implication is that universities should turn abstract cyber-governance language into visible workflows. For students, this might include registration-fee scam scenarios, scholarship-message verification, and learning-platform credential checks. For faculty, it might include manuscript, grading, data-sharing, and student-record impersonation scenarios. For administrative and professional staff, it might include procurement, payroll, HR, records, and vendor-payment verification. The core message should be shared across campus, while examples remain role-specific enough to be recognizable.

7.5. Role-Based Similarity at the Composite Level

The absence of statistically meaningful role differences is informative but should be read carefully. Neither the three-role ANOVAs nor the pooled student-versus-employee Welch analyses revealed significant differences across the focal composite constructs, and effect sizes were trivial. Broad legal, risk, trust, vigilance, and safety orientations may therefore be more similar than different across stakeholder roles in this sample.

The result does not, however, establish measurement invariance or show that role-specific risks are identical. Broad composites may be insensitive to differences in task exposure, message type, data sensitivity, or institutional authority. A campus-wide baseline intervention is therefore reasonable, but it should be supplemented with role-specific simulations and, in future work, measurement-invariance testing across students, faculty members, and professional staff.

7.6. Theoretical and Practical Contributions

The study makes three connected contributions. It positions legal literacy within digital-safety research as a compliance architecture that can strengthen trust without guaranteeing behavior. It specifies vigilance as the proximal mechanism through which trust becomes behaviorally useful. And it draws evidence from an underrepresented higher-education and UAE legal-governance context, showing that AI risk perception functions as more than a fear-oriented construct: it is associated with protective conduct when measured in terms of severity, vulnerability, and credibility amplification. These contributions should not be overstated as a discovery that human oversight is needed when AI is used, since that conclusion is already familiar in AI governance and cybersecurity. The more specific point is that, within these data, oversight becomes operationally meaningful when trust is translated into verification routines, reporting pathways, and incident-escalation behavior.

Practical interventions should be concrete. UAE universities can introduce legal-literacy micro-training based on cybercrime and personal-data-protection rules; embed AI-deception scenarios into orientation and staff development; require out-of-band verification for unusual credential, payment, data-transfer, or assessment requests; provide visible reporting buttons and incident-escalation flowcharts; publish response-time expectations; and normalize low-blame reporting so that uncertainty is escalated early. Such measures would operationalize trust-with-verification instead of relying on passive confidence in institutional-looking digital content.

8. Limitations and Future Research

Several limitations should guide interpretation. The design was cross-sectional, so the SEM pathways describe modeled associations and should not be read as causal effects. Although $N = 710$ is substantial for a university-stakeholder survey, the model spans many theoretically related dimensions, and the number of observations should be weighed against that dimensionality, the use of subscale indicators, and the non-probability nature of recruitment; the results are best treated as a cautious, model-based test rather than definitive validation of CALT-V across all UAE higher education. The sample was role-balanced but non-representative, and neither institutional coverage nor a response rate could be reported precisely because recruitment used decentralized, anonymous channels. Administering the questionnaire only in English may also have shaped who participated and how items were interpreted in a multilingual higher-education environment.

The study relied mainly on self-report, which may be affected by social desirability, optimistic self-assessment, and common method bias; procedural remedies were used, but no marker-variable or full item-level common-method-factor test was reported. The instrument combined adapted and study-specific scales, so while the CFA provides subscale-level

validation evidence, item-level alpha and omega coefficients, item wording, and further external validation should be supplied in Supplementary Materials and pursued in future work. Because the analysis used subscale indicators, the strong fit indices should not be taken as proof that the item-level structure is equally sound.

Objective legal knowledge was kept separate from the latent self-reported legal-literacy construct, and future research should test whether objective knowledge, perceived legal literacy, and procedural efficacy form a higher-order or bifactor structure. The final SEM involved limited, theory-guided respecification; although no post hoc error covariances were added merely to improve fit, alternative models should be compared in independent samples. Role-based comparisons were conducted at the composite level without measurement-invariance testing, and configural, metric, and scalar invariance should be established before stronger conclusions are drawn about stakeholder similarity.

Future research should use larger multi-institutional samples, item-level measurement models, behavioral simulations, phishing and AI-deception recognition tasks, incident-response exercises, and longitudinal training designs to determine whether legal-literacy micro-training and verification-routine interventions produce measurable gains in digital safety behavior. Comparative studies across emirates, institution types, and language contexts could also examine whether governance clarity moderates the translation of trust into vigilance.

9. Conclusions

This study reframes digital safety in UAE universities as a legally informed, trust-calibrated, and vigilance-dependent behavioral process. Across 710 university stakeholders, legal literacy, AI risk perception, trust, vigilance, and digital safety behaviors formed a coherent empirical system, yet the modeled pathway to safety was selective: AI risk perception and vigilance were most closely linked to protective behavior, while trust contributed mainly through vigilance.

The CALT-V account therefore offers both a theoretical and an applied contribution. Theoretically, it organizes legality, risk appraisal, trust, and vigilance within a single behavioral pathway and shows how these constructs relate in a UAE university sample. Practically, it provides a basis for university cyber-governance programs that combine UAE-specific legal literacy, AI-deception scenarios, out-of-band verification protocols, and low-friction reporting systems. Its central message is restrained but consequential for practice: digital safety depends less on trust alone than on calibrated trust enacted through vigilance. This message is compatible with the broader human-in-the-loop literature, but the study's specific contribution is to locate that principle within a particular legal, institutional, and higher-education setting. Because the evidence is cross-sectional, multidimensional, and non-representative, the conclusion should be treated as a theoretically informed association that requires replication, larger samples, and behavioral testing.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/socsci15090577/s1>. Supplementary File S1: Statistical analysis output and complete CALT-V survey instrument, including response formats, item codes, scoring instructions, descriptive statistics, reliability analyses, correlations, role-based comparisons, CFA results, and SEM direct, indirect, and total effects; Supplementary File S2: CALT-V scale/subscale-level dataset in Microsoft Excel format ($N = 710$), including the dataset and variable labels; Supplementary File S3: CALT-V reconstructed scale/subscale-level dataset in IBM SPSS Statistics format ($N = 710$).

Author Contributions: Conceptualization, A.S., M.A.M. and K.A.; methodology, A.S., M.A.M. and K.A.; formal analysis, A.S., M.A.M. and K.A.; investigation, A.S., M.A.M. and K.A.; data curation, A.S., M.A.M. and K.A.; validation, A.S., M.A.M. and K.A.; visualization, A.S., M.A.M. and K.A.; writing—original

draft preparation, A.S., M.A.M. and K.A.; writing—review and editing, A.S., M.A.M. and K.A., with M.A.M. taking primary responsibility for preparing the final version of the manuscript and completing all revisions required during the editorial and peer-review processes; project administration, A.S., M.A.M. and K.A.; funding acquisition, A.S. All authors contributed substantially and equally to the work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Ajman University Internal Research Grant No. 2025-IRG-COL-2.

Institutional Review Board Statement: The study was reviewed and approved by the Research Ethics Committee, Deanship of Graduate Studies and Research, Ajman University, UAE. The approval was granted on 5 November 2025, for the proposal titled *AI, Cybercrime, and Data Privacy in the UAE: Public Perceptions and Legal Awareness* under Project Number: L-F-H-4-Nov. The study was conducted in accordance with the approved research proposal, approved informed consent procedures, and approved research tools.

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

Data Availability Statement: The raw data supporting the conclusions of this article are provided as electronic Supplementary Files.

Acknowledgments: The authors gratefully acknowledge the support provided by the Ajman University Internal Research Grant No. 2025-IRG-COL-2. They also thank the students, faculty members, and administrative/professional staff who participated in the survey. The findings and conclusions expressed in this article are solely those of the authors and do not necessarily reflect the views of the funding institution.

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

Appendix A. Scale Architecture and Key Item Domains

This appendix summarizes the instrument domains used in the study. The full item wording and scoring key should be supplied as Supplementary Material.

Legal literacy. Rights literacy assesses perceived understanding of personal-data and cyber-safety rights in the university context. Obligations literacy assessed perceived understanding of lawful and responsible digital conduct. Procedural literacy assessed knowing how to verify, report, or escalate suspicious digital interactions. Objective legal knowledge consisted of six dichotomous items addressing UAE cybercrime and personal-data-protection principles.

AI risk perception. Items assessed perceived severity, perceived vulnerability, and AI-enabled credibility amplification, including the possibility that AI-generated text, synthetic voices/images, personalization, and automation can make deceptive university communications more convincing.

Trust. Institutional trust items assessed confidence in university communication, reporting channels, incident handling, and governance fairness. Technology/AI trust items assessed perceived reliability of digital and AI-mediated systems. The overreliance indicator captured passive confidence in institutional-looking or system-generated messages. It was reverse-scored for the descriptive overall trust total, while the CFA/SEM retained it as a reverse-oriented indicator under the original T_{over} label.

Vigilance. Normative vigilance items assessed endorsement of verification principles. Behavioral vigilance items assessed habits such as pausing before acting, checking sender identity, confirming unusual requests through a second channel, and reporting uncertain interactions.

Digital safety behaviors. Items covered secure authentication, cautious link and attachment handling, password/account hygiene, data minimization, verification of unusual requests, and reporting or escalation of suspected cyber incidents.

Governance conditions. Governance clarity items assessed whether policies, reporting channels, and cyber-safety responsibilities were understandable and visible. Enforcement items assessed perceptions that rules were consistently applied and that incidents generated appropriate institutional follow-up.

References

- Aggarwal, Gunjan, Aryan Kumar Pandey, and Swati Singal. 2025. Automated cyberattacks and social engineering. In *Digital Deception: Uncovering the Dark Side of AI in Social Networks*. Edited by Rupa Rani, Prashant Upadhyay, Rohit Sahu, Satya Prakash Yadav and Hardeo Kumar Thakur. Sharjah: Bentham Science Publishers, pp. 126–43. [\[CrossRef\]](#)
- Akdemir, Naci, and Christopher James Lawless. 2020. Exploring the human factor in cyber-enabled and cyber-dependent crime victimisation: A lifestyle routine activities approach. *Internet Research* 30: 1665–87. [\[CrossRef\]](#)
- Aregghan, Edoise. 2023. From data breaches to deepfakes: A comprehensive review of evolving cyber threats and online risk management. *Communication in Physical Sciences* 9: 538–53.
- Bird and Bird. 2021. How Does the New UAE Federal Decree Law on Personal Data Protection Compare Against the GDPR? December 15. Available online: <https://www.twobirds.com/en/insights/2021/uae/how-does-the-new-uae-federal-decree-law-on-personal-data-protection-compare-against-the-gdpr> (accessed on 26 July 2026).
- Boss, Scott R., Dennis F. Galletta, Paul Benjamin Lowry, Gregory D. Moody, and Peter Polak. 2015. What do systems users have to fear? Using fear appeals to engender threats and fear that motivate protective security behaviors. *MIS Quarterly* 39: 837–64. [\[CrossRef\]](#)
- Bullock, Justin B., Janet VT Pauketat, Hsini Huang, Yi-Fan Wang, and Jacy Reese Anthis. 2025. Public opinion and the rise of digital minds: Perceived risk, trust, and regulation support. *Public Performance & Management Review* Advance online publication. [\[CrossRef\]](#)
- Ewick, Patricia, and Susan S. Silbey. 1998. *The Common Place of Law: Stories from Everyday Life*. Chicago: University of Chicago Press.
- Ferrario, Andrea, Michele Loi, and Eleonora Viganò. 2020. In AI we trust incrementally: A multi-layer model of trust to analyze human–artificial intelligence interactions. *Philosophy & Technology* 33: 523–39. [\[CrossRef\]](#)
- Fornell, Claes, and David F. Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18: 39–50. [\[CrossRef\]](#)
- Habbal, Adib, Mohamed Khalif Ali, and Mustafa Ali Abuzaraida. 2024. Artificial intelligence trust, risk and security management (AI TRISM): Frameworks, applications, challenges and future research directions. *Expert Systems with Applications* 240: 122442. [\[CrossRef\]](#)
- Ifinedo, Princely. 2012. Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers & Security* 31: 83–95. [\[CrossRef\]](#)
- Islam, Ekramul, Carsten Rudolph, and Gillian Oliver. 2025. Managing cyber harm: A survey of challenges, practices, and opportunities. *Information Security Journal: A Global Perspective* 34: 424–54. [\[CrossRef\]](#)
- Jabir, Raja, John Le, and Chau Nguyen. 2025. Phishing attacks in the age of generative artificial intelligence: A systematic review of human factors. *AI* 6: 174. [\[CrossRef\]](#)
- Kaplan, Alexandra D., Theresa T. Kessler, J. Christopher Brill, and Peter A. Hancock. 2023. Trust in artificial intelligence: Meta-analytic findings. *Human Factors* 65: 337–59. [\[CrossRef\]](#) [\[PubMed\]](#)
- Klingbeil, Artur, Cassandra Grützner, and Philipp Schreck. 2024. Trust and reliance on AI: An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior* 160: 108352. [\[CrossRef\]](#)
- Laux, Johann, Sandra Wachter, and Brent Mittelstadt. 2023. Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and trust. *Regulation & Governance* 18: 3–32. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lee, John D., and Katrina A. See. 2004. Trust in automation: Designing for appropriate reliance. *Human Factors* 46: 50–80. [\[CrossRef\]](#) [\[PubMed\]](#)
- National Institute of Standards and Technology. 2023. *Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1)*. Washington, DC: U.S. Department of Commerce. [\[CrossRef\]](#)
- National Institute of Standards and Technology. 2024. *The NIST Cybersecurity Framework (CSF) 2.0 (NIST CSWP 29)*. Washington, DC: U.S. Department of Commerce. [\[CrossRef\]](#)
- Neyazi, Taberez Ahmed, Sheryl Wei Ting Ng, Mitchell Hobbs, and Audrey Yue. 2023. Understanding user interactions and perceptions of AI risk in Singapore. *Big Data & Society* 10: 20539517231213823. [\[CrossRef\]](#)
- Oerlemans, Jan-Jaap, Anne de Hingh, and Wytse van der Wagen. 2024. Types of cyber-enabled crime and their criminalisation. In *Essentials in Cybercrime: A Criminological Overview for Education and Practice*, 2nd ed. Edited by Wytse van der Wagen, Jan-Jaap Oerlemans and Marleen Weulen Kranenbarg. The Hague: Eleven International Publishing, pp. 97–143.
- Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee, and Nathan P. Podsakoff. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88: 879–903. [\[CrossRef\]](#) [\[PubMed\]](#)

- Rogers, Ronald W. 1975. A protection motivation theory of fear appeals and attitude change. *Journal of Psychology* 91: 93–114. [CrossRef] [PubMed]
- Schmitt, Marc, and Ivan Flechais. 2024. Digital deception: Generative artificial intelligence in social engineering and phishing. *Artificial Intelligence Review* 57: 324. [CrossRef]
- Stanton, Jeffrey M., Kathryn R. Stam, Paul Mastrangelo, and Jeffrey Jolton. 2005. Analysis of end user security behaviors. *Computers & Security* 24: 124–33. [CrossRef]
- Sulaiman, Noor Suhani, Muhammad Ashraf Fauzi, Walton Wider, Jegatheesan Rajadurai, Suhaidah Hussain, and Siti Aminah Harun. 2022. Cyber-information security compliance and violation behaviour in organisations: A systematic review. *Social Sciences* 11: 386. [CrossRef]
- Twomey, John, Didier Ching, Matthew Peter Aylett, Michael Quayle, Conor Linehan, and Gillian Murphy. 2023. Do deepfake videos undermine our epistemic trust? A thematic analysis of tweets that discuss deepfakes in the Russian invasion of Ukraine. *PLoS ONE* 18: e0291668. [CrossRef] [PubMed]
- United Arab Emirates. 2021a. Federal Decree-Law No. (34) of 2021 on Countering Rumours and Cybercrimes. UAE Legislation. Available online: <https://uaelegislation.gov.ae/en/legislations/1526/download> (accessed on 26 July 2026).
- United Arab Emirates. 2021b. Federal Decree-Law No. (45) of 2021 Concerning the Protection of Personal Data. UAE Legislation. Available online: <https://uaelegislation.gov.ae/en/legislations/1972/download> (accessed on 26 July 2026).
- Wong, Lai-Wan, Garry Wei-Han Tan, Keng-Boon Ooi, and Yogesh Dwivedi. 2024. The role of institutional and self in the formation of trust in artificial intelligence technologies. *Internet Research* 34: 343–70. [CrossRef]
- Zariski, Archie M. 2014. *Legal Literacy: An Introduction to Legal Studies*. Athabasca: Athabasca University Press. [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.