Journal of Computer and Communications, 2026, 14(3), 196-219
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

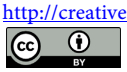
Predictors of Phishing Detection Confidence among Defense Engineering Professionals

Joseph Brickley, Kutub Thakur, Michael Krantz

Professional Security Studies Department of New Jersey City University, Jersey City, USA
Email: JBrickley@njcu.edu, KThakur@njcu.edu, MKrantz@njcu.edu

How to cite this paper: Brickley, J., Thakur, K. and Krantz, M. (2026) Predictors of Phishing Detection Confidence among Defense Engineering Professionals. *Journal of Computer and Communications*, **14**, 196-219.
https://doi.org/10.4236/jcc.2026.143010

Received: February 11, 2026
Accepted: March 27, 2026
Published: March 30, 2026



Abstract

Phishing is a persistent organizational threat, yet most empirical work emphasizes post-exposure susceptibility rather than the confidence judgments that shape behavior when suspicious messages arrive. This study examines predictors of phishing-detection confidence in a professional sample of U.S. defense-industry engineers (N = 97). Participants completed a validated survey capturing email comfort, perceived Internet proficiency, Self-Reported Phishing-Cue Knowledge, cue-checking frequency, use of sanctioned confirmation channels, and Big Five personality traits. Multiple linear models with standardized coefficients and multicollinearity checks estimate each factor's unique contribution. Three predictors show the largest and most consistent associations with detection confidence: email comfort, perceived Internet proficiency, and Self-Reported Phishing-Cue Knowledge. Emotional stability and openness contribute at smaller magnitudes, while cue-checking frequency and use of a confirmation channel exhibit secondary effects. Years of experience and recent training volume are not reliable drivers once core skills and knowledge are included. Findings support a confidence-mediated account of phishing decisions and suggest practical interventions: hands-on practice that improves email fluency and web problem-solving, explicit instruction on cue schemas, and low-friction confirmation workflows. Limitations include a single occupational context and reliance on retrospective confidence; implications for replication and program design are discussed.

Keywords

Confidence, Human Factors, Organizational Security, Phishing, Security Training, Self-Efficacy, Social Engineering, Usable Security

1. Introduction

Phishing remains one of the most persistent social-engineering threats in organi-

zations, repeatedly implicated in incidents and breaches [1] [2]. Much of the empirical literature emphasizes susceptibility after exposure, who clicks or replies and under what conditions, focusing on deceptive page design, social proof, urgency, and spoofing tactics [3]-[5]. That emphasis clarifies many determinants of being deceived but leaves less explained about the judgment users make before acting, how confident they feel in detecting a suspicious message. Confidence matters operationally: it influences whether an employee slows down, inspects surface and header cues, or seeks an authorized confirmation channel before executing a potentially risky action [3]-[5]. A clearer account of what drives phishing-detection confidence offers a direct lever for risk reduction because confidence can be shaped by training, process design, and tool support [6]-[8].

Building on this premise, we model detection confidence as a psychologically grounded belief state related to self-efficacy, the belief in one's capability to perform a task under typical constraints [6]. Prior usable-security work links self-efficacy with secure practice, suggesting that users who feel competent are more likely to enact recommended controls [7] [8]. Translating that insight to phishing, we examine drivers that are both learnable and actionable in enterprise programs: 1) technology comfort (comfort using email; perceived Internet proficiency), 2) knowledge of phishing cues captured by widely trained indicators (e.g., sender anomalies, link/attachment scrutiny), 3) frequency of cue checking and use of sanctioned confirmation channels (behavioral tendencies that organizations can standardize), and 4) personality traits (Big Five) previously associated with cautious responding or deception resistance [3]-[5] [7] [8]. We analyze these factors in a professional sample of U.S. defense-industry engineering staff using linear models with multicollinearity checks [9].

Research Question

RQ1: How do user characteristics affect users' retrospective confidence levels toward phishing detection?

RQ2: How do cues affect users' retrospective confidence levels?

Significance: By isolating rank-ordered, trainable predictors of detection confidence in a real workforce, this study provides a compact basis for prioritizing training content and operational safeguards in enterprise phishing programs.

2. Related Work

2.1. Phishing Deception and User Decision Processes

Phishing is defined here as the use of deceptive, computer-mediated communications to induce a security-relevant action, such as credential submission, payment authorization, attachment execution, or disclosure of sensitive information, by misrepresenting identity or context [10] [11]. As a subtype of social engineering, phishing combines technical and psychological techniques to elicit trust and exploit routine communication patterns. In organizational settings, attackers frequently seek to bypass established "trusted paths" (e.g., authenticated portals,

ticketing systems) by persuading users either to reply directly with information or to click through to an attacker-controlled site where the disclosure occurs [12] [13]. **Figure 1** illustrates the trusted-path concept and the attacker's goal of keeping users in the adversary's lane rather than returning to an authenticated channel. Within this work, deception cues denote surface, header, and contextual indicators (e.g., discrepancies between display and destination URLs, sender anomalies, unexpected attachments, urgency framing) that can support discrimination between benign and malicious messages.

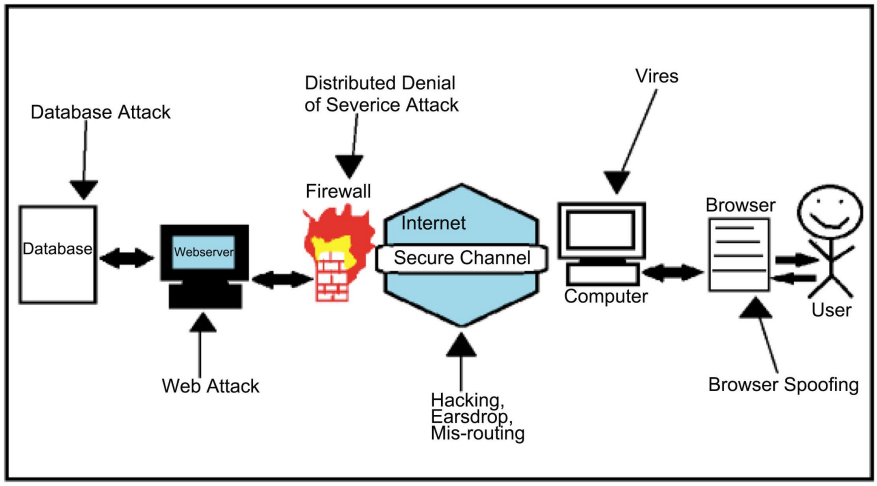


Figure 1. Trusted path redrawn (Adapted by Author from Li & Wu, 2003 [12]).

Industry analyses consistently situate phishing among the leading precursors to incidents and breaches, underscoring its operational impact on enterprises of all sizes [1] [2]. Foundational HCI/IS studies demonstrate how visual realism, identity/URL spoofing, and message framing systematically increase compliance, even among technically literate users who express high security awareness [3]-[5] [14] [15]. Small manipulations in domain strings and sender presentation, together with polished brand impersonation, can create an initial sense of legitimacy sufficient to trigger routine responding before analytic checking occurs [14] [15]. Content analyses further document the role of persuasion principles, especially authority and scarcity, in compressing deliberation windows and biasing toward rapid action (e.g., executive spoofing, deadline pressure, threat of negative consequences) [16]-[19]. Likeability/affinity and appeals to social proof also appear in attacker content, although their effects vary across studies and populations [16] [20]-[22]. As summarized in **Figure 2**, when mismatches between expectations and indicators are detected, users branch toward alternative actions rather than the phisher-suggested path [23].

Cognitive accounts of user decision making help explain these outcomes. Under time pressure and email load, users often default to peripheral cues (appearance, brand familiarity, sender label) rather than engaging in cue-analytic processing, consistent with dual-process and elaboration likelihood perspectives [3]

[5] [24]-[26]. Models tailored to online deception characterize detection as a multi-stage judgment: activation of suspicion upon encountering inconsistent cues; hypothesis generation that the content may be deceptive; hypothesis testing via additional evidence seeking (e.g., inspecting headers, hovering links, cross-channel verification); and global assessment culminating in action or avoidance [27]. Extensions to email contexts emphasize that priming (e.g., recent warnings or simulations) and individual differences influence whether relevant cue schemas are retrieved and applied at the moment of decision [28]. Building on this stream, the present study treats retrospective detection confidence as a proximal belief state that governs whether users slow down, interrogate cues, and seek confirmation before acting [9]. As shown in **Figure 3**, detection unfolds as a multi-stage judgment: activation, hypothesis generation, hypothesis testing, and global assessment.

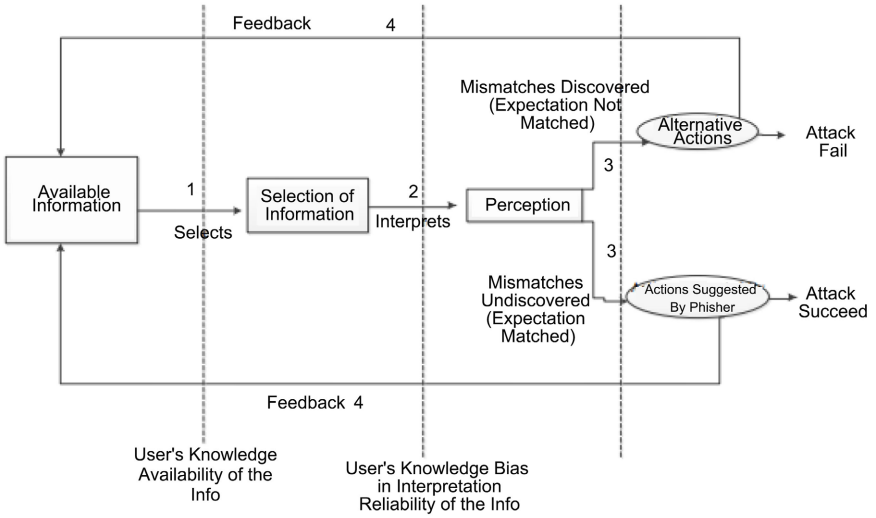


Figure 2. Email decision-making model (Adapted by Author from Xun *et al.*, 2008 [23]).

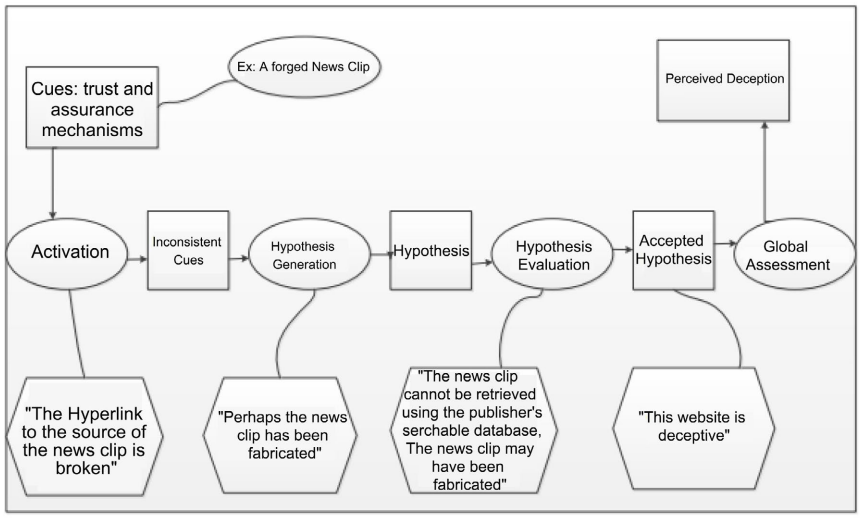


Figure 3. Multi-stage deception detection model (Adapted by Author from Grazioli, 2004 [27]).

Comparative work distinguishes detectors from victims by process as well as outcome. Detectors tend to notice inconsistent cues, test hypotheses, and select richer confirmation channels when uncertain; victims more often rely on appearance cues, misinterpret indicators, or choose weak confirmation paths [14] [23] [26] [26] [29] [30]. These patterns motivate the study's focus on trainable levers, technology comfort, cue knowledge, routinized cue-checking, and sanctioned confirmation pathways, and on confidence as the mechanism linking these levers to behavior within operational environments. As shown in **Figure 4**, risky in-thread replies keep users in the adversary's lane, whereas safer out-of-thread confirmation returns the user to a trusted path, motivating our emphasis on sanctioned channels.

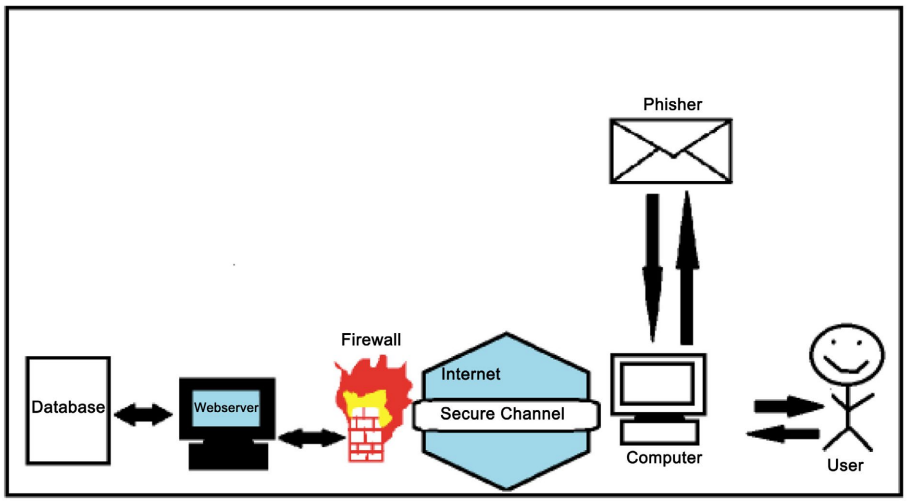


Figure 4. Phishing reply flow (Adapted by Author from Alseadoon, 2014 [14]).

2.2. Confidence as Self-Efficacy in Usable Security

In this study, we treat detection confidence as a task-focused belief state aligned with self-efficacy, one's perceived capability to execute the behaviors required to achieve a specific outcome under typical constraints [6]. Within security practice, higher self-efficacy has been associated with stronger adherence to recommended behaviors (e.g., checking indicators, using sanctioned channels) and reduced intention to engage in unsafe acts [7] [8]. Framed this way, phishing detection is not only about knowledge of cues but also about the belief that one can apply that knowledge efficiently in routine work conditions (email load, deadlines, social pressure).

Prior work distinguishes prospective and retrospective forms of confidence. Prospective confidence (before acting) reflects beliefs about one's capability and overlaps conceptually with self-efficacy [6] [31]-[35]. Retrospective (judgmental) confidence is assessed after a decision and reflects the perceived correctness of that decision [36] [37]. Behavioral decision research links both forms to downstream behavior: higher confidence generally predicts more decisive action, whereas lower confidence predicts additional information search and verification

[34]. In phishing contexts, elevated confidence, especially when outpacing actual skill, has been tied to faster responding, reduced cue inspection, and greater risk-taking [32] [35]. Conversely, lower confidence often triggers confirmation via richer channels (e.g., calling a known contact or opening a ticket) prior to action [22] [33].

A second stream emphasizes where confidence comes from. Social cognitive theory posits four primary sources of efficacy beliefs: mastery experiences, vicarious learning, verbal persuasion, and physiological states [31]. In enterprise programs, these map directly to hands-on simulation and feedback (mastery), modeling via worked examples, explicit policy/process messaging, and fatigue-aware delivery. Consistent with this account, studies show that training may raise security self-efficacy and improve certain protective behaviors, although gains are uneven and can be offset by overconfidence when task demands or cue complexity are high [7] [8] [32]. Evidence from printed-email and lab paradigms further indicates that participants can report high confidence while misclassifying phish at meaningful rates, highlighting the risk of miscalibrated confidence [3] [5] [32].

Finally, research on confidence calibration suggests that motivational self-enhancement and perceived competence in a domain can inflate confidence beyond performance, producing systematic errors under uncertainty [32] [35]. In email triage, brand familiarity and routine exposure may create an "illusion of validity," encouraging peripheral processing of polished impostors and premature action [3] [25] [36] [37]. Together, these strands motivate our focus on trainable levers (email comfort, Internet problem-solving, cue knowledge) and routinized safeguards (cue-checking, sanctioned confirmation pathways) as mechanisms for improving confidence and calibration in operational settings, with retrospective detection confidence analyzed as the proximal judgment that links these levers to behavior [9] [22] [32] [33].

2.3. Technology Comfort and Cue Knowledge

In practice, detection improves when users are comfortable with the medium and know exactly what to look for. Consistent with prior work, technology comfort, routine facility with mail clients, and basic web problem-solving matter most when inbox load and time pressure push users toward peripheral judgments [3] [5] [25]. Simply handling more messages is not protective; higher volume can increase haste and appearance-based responding. What differentiates detectors in operational settings is that comfort translates into habitual indicator inspection before action (sender anomalies, display vs. destination URL, unexpected attachments) rather than relying on visual polish or brand familiarity [13] [25].

Comfort alone is insufficient without cue knowledge. Foundational HCI/IS studies show that small URL/domain manipulations, spoofed sender presentation, and polished brand impersonation reliably elicit compliance unless users actively interrogate indicators [3] [14] [30]. Eye-tracking and decision-process work further show detectors attend to diagnostic regions (headers, links), while victims

overemphasize global appearance—precisely the failure mode attackers design for [26] [29]. Prior findings also distinguish knowing that "phishing exists" from knowing which cues distinguish legitimate from deceptive messages at decision time; the latter predicts avoidance, the former does not [13] [19] [25] [27].

Tool assistance helps but does not remove the need for internalized schemas. Filters and toolbars reduce some threats, yet are inconsistently heeded and can be bypassed by high-fidelity spoofs, leaving human indicator checking decisive at the moment of choice [3] [14] [24] [30]. Training that explicitly rehearses cue identification and verification (hover links, inspect headers, and when uncertainty persists, use a sanctioned confirmation path) improves discrimination for many users, though gains decay without reinforcement; critically, some users retain high confidence while misclassifying sophisticated phish, underscoring the need to calibrate confidence, not merely increase it [7] [8] [22] [24].

Consistent with the trusted-path framework described above, two organizational levers are particularly salient: 1) increasing email fluency and web problem-solving capability so that diagnostic indicators are accessed efficiently under workload constraints, and 2) making cue schemas explicit and sufficiently rehearsed to support automatic retrieval at the point of decision. When uncertainty persists, routinized use of sanctioned confirmation channels shifts users away from in-thread replies—an attacker-favored pathway—toward independently sourced verification, a pattern repeatedly observed among detectors in prior work [13] [22] [23] [25].

2.4. Behavioral Safeguards

When a suspicious message bypasses technical controls, safer decision pathways involve exiting the attacker-controlled communication thread and returning to an independently authenticated channel. Prior work characterizes initial triage as a deliberate inspection of diagnostic indicators rather than reliance on visual polish or brand familiarity. Diagnostic behaviors documented in the literature include expanding headers, comparing display names to underlying sender addresses, examining display versus destination URLs, and treating unexpected attachments or urgency framing as potential risk signals. Empirical studies show that minor domain manipulations and high-fidelity visual impersonation reliably elicit compliance unless users slow processing and interrogate indicators [3] [14] [30]. Eye-tracking and decision-process research further demonstrate that detectors allocate attention to diagnostic regions (e.g., headers and links), whereas victims disproportionately attend to global appearance cues [27] [29].

When uncertainty persists, confirmation behavior becomes critical. Research consistently distinguishes in-thread confirmation—replying within the same email chain or clicking embedded links—from independently sourced verification. The latter involves switching media or channel (e.g., opening a ticket, calling a directory-listed extension, or contacting the purported sender through an independently obtained address) and restores what prior models describe as a trusted path [13].

Within Grazioli's multi-stage framework, this behavior reflects the hypothesis-testing phase, where outcomes depend on both the quality of evidence sought and the richness of the channel used to obtain it [27] [28].

Effective verification addresses both identity validation and alignment with established organizational processes before the execution of sensitive actions such as credential submission, payment authorization, or data disclosure. Studies indicate that structured indicator inspection combined with out-of-thread confirmation reduces deception success rates, particularly under workload conditions [12] [27]. Checklist-based reinforcement and embedded cues within mail clients have been shown to improve retrieval of verification steps under time pressure and may mitigate overconfidence effects observed with high-fidelity spoofs [22] [32].

Tool assistance reduces exposure but does not eliminate the need for user judgment. Security toolbars and filters decrease threat prevalence; however, users frequently overlook or discount visual indicators, particularly when message realism is high [3] [14] [30]. Across studies, detectors consistently combine indicator inspection with sanctioned confirmation channels, whereas victims are more likely to reply in-thread or act without independent verification [13] [23] [27] [29].

2.5. Personality Factors

Personality differences shape how people approach uncertainty and social requests, so it is unsurprising that Big Five traits show modest but systematic associations with phishing judgments [5] [38]. Conscientiousness and emotional stability are typically linked to more deliberate, cue-analytic processing (checking headers/URLs, pausing under time pressure), whereas agreeableness, extraversion, and openness can, in certain framings, increase willingness to engage or comply (e.g., helpfulness to "urgent" requests, social reciprocity, curiosity to "learn more") [5] [38]. These effects are consistently smaller than skill/knowledge effects and can invert when urgency, authority, or workload pushes users toward peripheral processing [25].

These findings align with prior work indicating that personality traits exert smaller and more context-dependent effects on phishing judgments relative to task-relevant skills and knowledge. Emotional stability (low neuroticism) and openness contribute at secondary magnitudes to detection confidence once core, trainable levers—email comfort, perceived Internet proficiency, and Self-Reported Phishing-Cue Knowledge—are included [22] [25]. Mechanistically, dispositional calm may support hypothesis testing following suspicion activation, whereas openness may either facilitate cue exploration or, under high-fidelity impersonation, increase exploratory interaction when diagnostic schemas are not salient [13] [25].

Overall, personality appears to influence processing tendencies under uncertainty but accounts for substantially less variance in detection confidence than capability- and knowledge-based predictors. Consistent with the broader literature, dispositional traits shape how users approach ambiguous requests, whereas fluency and cue knowledge more directly determine confidence judgments in oper-

ational email contexts [5] [25] [38].

Ethical considerations also arise when incorporating personality insights into organizational security design. Prior research cautions against profiling or individualized punitive controls and instead emphasizes the importance of universal safeguards that support consistent protective behavior across user types [13] [22]. In this context, environmental and workflow-based interventions are positioned as structural supports that mitigate dispositional risk without targeting individuals directly.

2.6. Gap and Approach

Despite extensive work on phishing deception and user decisions, three gaps remain salient. First, most studies emphasize post-exposure susceptibility or printed/lab judgments and give limited attention to the belief state that governs behavior at the moment of choice—retrospective detection confidence (RDC) [3] [5] [32]. Related streams document cue effects and dual-process dynamics but typically stop short of modeling confidence as the proximal mechanism that determines whether users slow down, interrogate indicators, or seek confirmation [25]-[27]. Second, confirmation behavior is often discussed qualitatively but rarely operationalized as a measurable safeguard (channel choice/time-to-verification) in concert with cue use; yet detectors in field contexts reliably exit the attacker's lane and switch media when uncertain [13] [22] [27]. Third, many findings derive from student or convenience samples and single-cue or single-principle stimuli, limiting external validity for enterprise inboxes where workload, polished impersonation, and organizational workflow shape decisions [3] [22].

This study addresses these gaps directly. Detection confidence is treated as a task-focused belief state aligned with self-efficacy and is modeled as the proximal driver of triage behavior under routine constraints (load, deadlines, social pressure) [6]-[8] [32]. Cue knowledge is paired with routinized behavioral safeguards by measuring both indicators use (e.g., header and URL checks) and sanctioned confirmation channel use, linking them to confidence rather than only to accuracy [13] [22] [27]. To strengthen ecological validity, the sample comprises a professional U.S. defense-industry cohort (N = 97) whose daily work involves high-stakes communications and mandatory training, rather than lab volunteers evaluating static printouts [1]-[3].

The predictor set is compact and explicitly trainable: technology comfort (email comfort, perceived Internet proficiency), Self-Reported Phishing-Cue Knowled ge, cue-checking frequency, sanctioned confirmation use, and Big Five traits as secondary dispositions [5] [13] [22] [25] [39]. Standardized OLS models with block entry (first skills and knowledge, then behaviors, then personality) are estimated; multicollinearity checks and robustness diagnostics are reported, and interpretation emphasizes rank-ordered contributions to confidence rather than omnibus fit alone [9]. The design maps cleanly to enterprise levers (training content, workflow nudges, tool cues) and avoids over-reliance on demographic mod-

erators not used in the models.

Conceptually, the paper advances a confidence-mediated account of phishing decisions that integrates cue schemas and confirmation behavior within established deception models [22] [25] [27]. Practically, it delivers a prioritized, actionable set of drivers—email fluency, Internet problem-solving, and cue knowledge—that explains the bulk of variance in detection confidence, with personality playing a smaller role. Program guidance follows directly: teach indicators, rehearse out-of-thread verification, and instrument workflows to make the safer action the easier action in real inboxes [13] [22] [30].

3. Methods

3.1. Participants and Context

The sample comprised N = 97 full-time employees of a U.S. Department of Defense (DoD) contractor drawn from the firm's Lynchburg, Virginia division. This unit is predominantly technical: respondents included engineers and systems architects from associate through senior levels whose daily work involves high-stakes communication, change control, and adherence to formal security workflows. The organization conducts periodic phishing simulations and mandatory cybersecurity training; recruitment for the present study followed immediately after one such simulation featuring an authority-themed ACH/EFT direct-deposit confirmation email. Positioning the survey at this moment ensured that judgments referenced a concrete, recent inbox event rather than abstract scenarios, improving ecological validity relative to typical student/convenience samples reported in the literature [1]-[3] [13] [25]. Data were collected online in Summer 2022. Participation was voluntary and anonymous, and all outreach occurred after the employer's exercise concluded to minimize any perception of employment pressure. While access relied on convenience sampling, the professional, technical cohort aligns the study context with enterprise environments emphasized by organizational reporting and usable-security research [1]-[3] [13] [22] [25].

Sampling Transparency. The survey was distributed to 114 engineering employees within the selected division. Of these, 98 agreed to participate (85.96% response rate), and 97 completed responses were retained for analysis. Participation was voluntary and anonymous. Inclusion criteria consisted of full-time employees within the targeted engineering division; employees under the age of 18 were excluded. A convenience sampling approach was used due to the researcher's access to the organization. Only completed surveys were included in the final analytic dataset; no imputation procedures were applied.

3.2. Measures

All measures were administered online using 5-point Likert-type scales unless noted. Items were written in plain, work-proximal language and anchored to email triage or web problem-solving to maximize face validity for a professional audience. Multi-item indices were computed as means so higher scores reflect

"more" of the construct. Scale quality met accepted standards for social-science work: the administered battery yielded post hoc $\alpha = 0.753$, and construct-level reliability for composites is reported with descriptives in **Table 1** [31] [39].

Outcome (Retrospective Detection Confidence; Frame Variants). The primary outcome was retrospective detection confidence (RDC), defined as respondents' judgmental confidence regarding phishing-detection decisions. RDC was operationalized in four analytic "frames" to distinguish stimulus-anchored confidence from generalized detection confidence. The Specific Email frame (stimulus-anchored RDC) was computed as the mean of four items (Q26, Q28, Q29, Q31; 1 - 5 scale) referencing the embedded organizational phishing simulation email (e.g., "Upon receiving the email below, I was confident in my decision process"). This four-item composite demonstrated acceptable internal consistency ($\alpha = 0.753$).

In addition, three single-item generalized confidence indicators were analyzed separately: confidence detecting phishing emails, confidence detecting deceitful emails, and confidence detecting cybercrime-related emails (each 1 - 5 scale). These items were modeled independently and are referred to as the Phishing, Deceitful, and Cybercrime frames in the Results. For multi-item composites, scores were computed as arithmetic means. Higher values reflect greater retrospective detection confidence. No reverse coding was required.

Primary trainable predictors (skills and knowledge). Consistent with the study's emphasis on actionable levers, we captured 1) Email Comfort via a single face-valid item ("I am comfortable using email," Q30; 1 - 5); 2) Perceived Internet Proficiency ("How proficient are you at using the internet?" Q14; 1 - 5); 3) Self-Reported Phishing-Cue Knowledge ("I am knowledgeable about phishing emails," Q15; 1 - 5); and 4) Cue-Checking Frequency ("I inspect emails for phishing cues…," Q25; 1 - 5, from All the time to Never). The cue knowledge item reflects perceived knowledge of diagnostic phishing indicators rather than objective performance on cue-identification tasks. Together, these variables operationalize self-perceived capability and routine behaviors commonly associated with analytic processing and accurate discrimination in prior work [3] [5] [7] [8] [14] [25] [30].

Behavioral safeguards (confirmation). To index routinized verification, the survey assessed the likelihood of using sanctioned confirmation channels when uncertain ("How likely are you to ask someone for their opinion…," Q21; 1 - 5), followed by channel selection (face-to-face, telephone, IM, email; Q22), confidence in the consulted party (Q23), and influence on the final decision (Q24). The likelihood item (Q21) captures the stated intent to use a sanctioned confirmation channel. However, the primary safeguard predictor entered in the main models is Q24 ("The consulted opinion impacted my decision"), as it reflects enacted confirmation influence rather than stated likelihood. Channel-selection details (Q22 - Q23) support descriptives and robustness checks consistent with confirmation constructs in Grazioli and Alseadoon [13] [23] [27] [28].

Personality (dispositional controls). Given modest but reliable links between

disposition and phishing responses, we included single-item Big Five indicators adapted for brevity in an enterprise setting: Emotional Stability ("I remain calm in stressful situations," Q1), Agreeableness (Q2), Extraversion (Q3), Openness (Q4), and Conscientiousness (Q5), each on 1 - 5 scales [5] [25] [39] [40]. Personality is modeled as secondary to trainable skills/knowledge but provides controls for cautiousness versus engagement tendencies.

Experience and exposure (contextual controls). Contextual variables recorded years of email use (Q7), email load as typical daily volume (Q9), years of internet experience (Q13), and years of work experience (Q16), all in ordered categories. To represent training/priming, we captured counts of social-engineering, cybersecurity, and phishing-specific trainings (Q6, Q10, Q11; 1 - 5 each) plus training frequency (Q12). The three count items are averaged into a Training Volume index used in robustness checks and descriptives [7] [8] [17] [22]. Prior exposure was measured with self-victimization frequency (Q17) and social exposure (Q18 - Q19), given evidence that such experiences shift cue attention and confidence calibration [22] [25] [32].

Cue-process indicators (SRQ2). Finally, to characterize the immediate decision process for the embedded stimulus, respondents indicated which elements they checked (sender address, subject, body, hyperlink, none; Q32), the order of inspection (Q33), and the action after the first check (inspect more, delete, ignore, report, respond; Q34). These items map onto activation towards hypothesis generation/testing sequences emphasized in deception models and email-specific adaptations [22] [27] [28]. For parsimony in the main OLS models, cue-checking frequency (Q25) is the process variable entered; SRQ2 items inform process descriptives and sensitivity analyses.

Because cue knowledge was measured via self-assessment rather than objective cue-identification testing, estimates may reflect perceived competence rather than demonstrated discrimination accuracy. Self-reported capability measures are subject to common-method variance and potential overconfidence bias, particularly in security contexts where respondents may overestimate their detection skills [32] [34]. Accordingly, findings should be interpreted as associations between perceived cue knowledge and confidence rather than objective performance.

Scoring and Coding. All Likert-type items were coded on 1 - 5 scales. For most constructs, higher values reflect greater levels of the construct, including greater email comfort, greater cue-checking frequency, and greater retrospective detection confidence. Interpretation of descriptive statistics and model coefficients follows this general directionality unless otherwise noted.

Perceived Internet Proficiency (Q14) was coded 1 = Very proficient to 5 = Very un-proficient; therefore, lower numeric values indicate higher proficiency. Cue-Checking Frequency (Q25) was coded 1 = All the time to 5 = Never, such that lower values represent more frequent cue inspection. Confirmation Likelihood (Q21) was coded 1 = Very likely to 5 = Very unlikely, meaning lower values reflect a greater likelihood of using a sanctioned confirmation channel. Correlation signs

and regression coefficients reflect these coding directions.

Multi-item composites, including the stimulus-anchored RDC and Training Volume indices, were computed as arithmetic means of their constituent items. No items were reverse scored. Internal consistency estimates and zero-order intercorrelations are reported in **Table 1**. The modeling strategy (standardized OLS with block entry; multicollinearity and robustness checks) is described in Section 3.4 and implemented in Results [5] [9] [25] [32].

3.3. Procedures

Data collection was conducted in Summer 2022 via a short online survey administered a few days after the organization's routine phishing simulation concluded. This timing anchored respondents' judgments in a realistic, recent inbox event while clearly separating the research from the employer's internal exercise to avoid any perception of required participation. The email invitation described the study purpose and linked to an information/consent page outlining risks, benefits, voluntariness, and data handling; only after affirmative consent could participants proceed.

The instrument followed a fixed sequence to minimize construct carryover: user characteristics and experience (email, Internet, work), training/priming (counts and frequency), self-reported phishing-cue knowledge and cue-checking frequency, likelihood of using sanctioned confirmation channels with brief follow-ups on channel type and influence, personality indicators, and finally the retrospective detection confidence items anchored to the organization's example email. Items used neutral wording and concrete examples to reduce demand characteristics; the embedded email served solely as a shared reference point for the confidence judgments.

Participation was voluntary and uncompensated. The survey platform limited responses to one per participant, captured no personally identifying information, and did not retain IP addresses or device fingerprints. Responses were automatically coded and exported to an analysis dataset with anonymous case identifiers. Raw data access was restricted to the research team; de-identified data are stored on encrypted media with access controls and are retained for at least five years under institutional policy. The study protocol received ethics approval from the New Jersey City University Institutional Review Board, and all procedures conformed to applicable human-subjects protections.

3.4. Analysis Plan

Analyses were pre-specified to estimate the unique contribution of actionable factors to retrospective detection confidence (RDC) and to present effect sizes on a common scale. Ordinary least squares models were estimated with standardized coefficients so that magnitudes are directly comparable across predictors. Predictors were block-entered to reflect the theoretical ordering used throughout this work and the practical emphasis on trainable levers: Block 1 included

skills/knowledge variables (Email Comfort, Perceived Internet Proficiency, Self-Reported Phishing-Cue Knowledge, Cue-Checking Frequency); Block 2 added the behavioral safeguard, operationalized as confirmation influence (Q24: "the consulted opinion impacted my decision"), reflecting enacted confirmation rather than stated likelihood; Block 3 added dispositional controls (single-item Big Five indicators). Contextual covariates (years of email use, daily email volume, years of Internet use, years of work experience) and a Training Volume index (mean of social engineering, cybersecurity, and phishing-specific training counts) were summarized descriptively and then introduced in robustness specifications to verify that the rank ordering of the core effects was not an artifact of omitted experience or training variance.

Composite scores were computed as means of their constituent items (e.g., RDC from Q26, Q28, Q29, Q31). Unless otherwise noted, higher scores reflect more of the construct. Descriptive statistics (means and standard deviations), internal consistency for multi-item composites (post hoc $\alpha = 0.753$), and zero-order correlations are reported with the Results to provide a transparent view of scale quality and associations prior to modeling.

Model diagnostics were conducted for every specification. Multicollinearity was assessed via variance inflation factors (target VIF < 5), with condition indices inspected when VIFs approached common thresholds. Residual diagnostics included component-plus-residual plots for linearity and additivity, Q-Q plots for normality, and Breusch–Pagan/White tests for heteroskedasticity. Where heteroskedasticity was indicated, heteroskedasticity-consistent standard errors (HC3) were reported without altering point estimates. Influence and stability were evaluated using leverage, studentized residuals, and Cook's distance; sensitivity analyses re-estimated models after excluding any cases exceeding conventional influence cutoffs to confirm that substantive conclusions were unchanged.

Because item-level missingness was minimal, listwise deletion was used for the main models. A sensitivity check using mean-imputed composites produced substantively identical results and is omitted for brevity. Additional sensitivity analyses 1) dichotomized the confirmation likelihood at "Likely/Very likely" and re-estimated the Block 2 effect, and 2) added the Training Volume index and contextual covariates as a final block; in both cases, the ordering and interpretation of the core skills/knowledge and safeguard effects remained stable.

Main findings are presented as standardized coefficients with 95% confidence intervals. Complete item wording, coding rules, alternative model specifications (e.g., training-volume substitutions, exclusion of single-item predictors), and full coefficient tables with diagnostics are available from the authors upon reasonable request. This completes the Method. The next section proceeds to the empirical results, first the descriptives and reliability, then the block-entered OLS models and robustness checks, followed by a Discussion that interprets the rank-ordered predictors for theory and program design.

4. Results

RDC was computed as a four-item mean composite and showed acceptable internal consistency ($\alpha = 0.753$; Table 1). Descriptively, the sample reported high email comfort (M ≈ 1.4) and high perceived Internet proficiency (M ≈ 1.5), alongside frequent cue inspection (M ≈ 1.54). Zero-order associations with RDC concentrated in three trainable factors (Table 1): Email Comfort correlated positively across frames (r ≈ 0.38 - 0.50); Perceived Internet Proficiency tracked with RDC (scale coded 1 = very proficient to 5 = very un-proficient; r ≈ −0.40 for the single-stimulus frame and ≈ 0.41 - 0.49 in the other frames as reported); and Self-Reported Phishing-Cue Knowledge showed the largest positive bivariate relations for phishing, deceitful email, and cybercrime (r ≈ 0.60, 0.50, 0.55). Cue-checking frequency was small and frame-dependent (n.s. for the single-stimulus frame; r ≈ 0.21 - 0.24 otherwise). Confirmation items were weakly positive in bivariate terms, with "the consulted opinion impacted my decision" the strongest of the three (r ≈ 0.22 - 0.36). Personality indicators were secondary: Emotional Stability and Openness correlated weakly and positively with RDC; Extraversion showed a single positive relation in the specific-email frame; Agreeableness and Conscientiousness were near zero. Years-based experience measures (email, Internet, work) were negligible or inconsistent.

Table 1. Descriptives, reliability, and zero-order correlations with RDC (by frame).

Measure	Mean	α/Items	r (RDC: Specific Email)	r (RDC: Phishing)	r (RDC: Deceitful)	r (RDC: Cybercrime)
Retrospective Detection Confidence (RDC)	1.90	$\alpha = 0.753/4$				
Email Comfort	1.40	—	0.384	0.487	0.493	0.498
Perceived Internet Proficiency*	1.50	—	−0.404	0.485	0.460	0.411
Self-Reported Phishing-Cue Knowledge	—	—	—	0.600	0.496	0.547
Cue-Checking Frequency	1.54	—	n.s.	0.210	n.s.	0.238
Confirmation: Opinion Impacted Decision	2.10	—	0.358	0.220	0.216	0.357
Personality: Emotional Stability	—	—	—	0.348	0.288	0.319
Personality: Openness	—	—	—	0.221	0.227	0.313
Personality: Extraversion	—	—	0.246	—	—	—
Personality: Agreeableness	—	—	—	—	—	0.194

Notes. SD intentionally omitted to keep the main table compact; α shown for RDC composite ($\alpha = 0.753$). Perceived Internet Proficiency coded 1 = very proficient to 5 = very unproficient; signs reflect coding. Detailed descriptives and diagnostics are available from the authors upon request.

Block-entered OLS models quantified unique contributions by predictor family (Table 2). In the skills/knowledge block, Email Comfort was a consistent, sizeable

positive predictor: specific email $\beta \approx 0.677$, phishing $\beta \approx 0.782$, deceitful $\beta \approx 0.676$, cybercrime $\beta \approx 0.899$. Perceived Internet Proficiency showed similarly robust positive effects across frames ($\beta \approx 0.530, 0.596, 0.514, 0.507$). Self-Reported Phishing-Cue Knowledge, operationalized within the priming family, was the dominant content variable for phishing, deceitful email, and cybercrime ($\beta \approx 0.653, 0.512, 0.659$). Cue-checking frequency contributed small, frame-specific increments (n.s. for the single-stimulus frame; phishing $\beta \approx 0.217$; deceitful n.s.; cybercrime $\beta \approx 0.247$). In the behavioral-safeguard family, only "the consulted opinion impacted my decision" reached significance, and only for the specific-email frame ($\beta \approx 0.426$); confirmation was n.s. elsewhere. Within personality, Emotional Stability and Openness produced small positive coefficients in several specifications (e.g., deceitful and cybercrime for Emotional Stability; phishing and cybercrime for Openness). Extraversion was isolated to the specific-email frame, and Agreeableness/Conscientiousness remained near zero. Model fit was low to moderate and aligned with expectations about proximal, trainable determinants of confidence: personality-only $R^2 \approx 0.084/0.168/0.123/0.204$; email-experience $R^2 \approx 0.160/0.249/0.294/0.263$; Internet-experience $R^2 \approx 0.165/0.257/0.218/0.183$; priming (including cue knowledge) $R^2 \approx 0.091/0.367/0.280/0.323$; confirmation $R^2 \approx 0.151/0.081/0.115/0.166$; cues $R^2 \approx 0.015/0.044/0.018/0.057$ (see Table 2 for the consolidated summary; full coefficient matrices and diagnostics are available from the authors upon request).

Table 2. Consolidated OLS summary by predictor family (coefficients as reported; frame: Specific/Phishing/Deceitful/Cybercrime).

Predictor Family/Term	Coeff. (Specific)	Coeff. (Phishing)	Coeff. (Deceitful)	Coeff. (Cybercrime)
Skills & Knowledge → Email Comfort	0.677 ($p < 0.001$)	0.782 ($p < 0.001$)	0.676 ($p < 0.001$)	0.899 ($p < 0.001$)
Skills & Knowledge → Perceived Internet Proficiency	0.530 ($p < 0.001$)	0.596 ($p < 0.001$)	0.514 ($p < 0.001$)	0.507 ($p < 0.001$)
Skills & Knowledge → Self-Reported Phishing-Cue Knowledge	—	0.653 ($p < 0.001$)	0.512 ($p < 0.001$)	0.659 ($p < 0.001$)
Skills & Knowledge → Cue-Checking Frequency	n.s.	0.217 ($p = 0.040$)	n.s.	0.247 ($p = 0.020$)
Behavioral Safeguard → Confirmation: Opinion Impacted Decision	0.426 ($p = 0.023$)	n.s.	n.s.	n.s.
Personality → Emotional Stability	n.s.	n.s.	0.231 ($p = 0.018$)	0.238 ($p = 0.001$)
Personality → Openness	n.s.	0.267 ($p = 0.048$)	n.s.	0.379 ($p = 0.001$)
Personality → Extraversion	0.266 ($p = 0.025$)	n.s.	n.s.	n.s.
Personality → Agreeableness/Conscientiousness	n.s.	n.s.	n.s.	n.s.
Model R^2 → Personality block	0.084	0.168	0.123	0.204
Model R^2 → Email experience block	0.160	0.249	0.294	0.263
Model R^2 → Internet experience block	0.165	0.257	0.218	0.183
Model R^2 → Confirmation block	0.151	0.081	0.115	0.166
Model R^2 → Priming (incl. Cue Knowledge)	—	0.367	0.280	0.323

Notes. Entries reproduce coefficients/R^2 already reported in Chapter 4. Cells marked "—" indicate not estimated in that frame. Full coefficient matrices and diagnostics are available from the authors upon request.

Robustness checks supported these conclusions. Multicollinearity was acceptable (typical VIFs ≈ 1.0 - 2.0; all <5), and Email Comfort and Perceived Internet Proficiency were empirically separable despite conceptual proximity. Alternative coding (e.g., dichotomizing confirmation as Likely/Very likely vs. other) did not change interpretation. Adding contextual covariates (years of email use, daily email volume, years of Internet use, years of work experience) and a Training-Volume index left the signs and relative magnitudes of the core skills/knowledge effects intact; in contrast, generic training volume was weakly or negatively related to RDC in bivariate views while cue knowledge remained positive. Assumption checks (linearity/additivity via component-plus-residuals; normality via Q-Q) were acceptable; HC3 standard errors addressed occasional heteroskedasticity without affecting point estimates. Influence statistics (leverage, studentized residuals, Cook's distance) revealed no cases that altered inference; exclusions beyond conventional thresholds left magnitudes and signs stable. Missingness was minimal; mean-imputed composites reproduced listwise-deletion results; detailed diagnostics are available from the authors upon request.

Substantively, RDC is primarily a function of trainable capability and knowledge rather than tenure or disposition. Email fluency and general Internet problem-solving, together with explicit knowledge of diagnostic phishing cues, account for most of the predictable variance; habitual cue-checking adds a smaller, frame-specific increment; sanctioned confirmation appears behaviorally protective but is not a strong determinant of reported confidence; personality contributes only secondary variance. These results justify programs that prioritize hands-on practice to increase email fluency and web problem-solving, explicit instruction and rehearsal of cue schemas, and low-friction out-of-thread confirmation pathways, while de-emphasizing years-based tenure as a proxy for confidence-relevant skill.

5. Discussion and Implications

The core finding is that retrospective detection confidence (RDC) is chiefly a function of trainable capability and knowledge rather than tenure or broad exposure. Mathematically, this is evidenced by large, standardized coefficients (β) for three predictors across frames: Email Comfort, Perceived Internet Proficiency, and Self-Reported Phishing-Cue Knowledge, with β estimates commonly in the 0.50 - 0.90 range (**Table 2**) and correspondingly moderate R^2 at the family level (**Table 2**) despite parsimonious models. These results indicate that detection confidence is most strongly associated with fluency in email tools, perceived Internet problem-solving capability, and explicit knowledge of diagnostic phishing cues. The pattern suggests that transferable skill and cue-schema mastery, rather than accumulated tenure, are the primary determinants of confidence in operational phishing judgments.

The skills/knowledge block consistently dominated. For example, Email Comfort and Internet Proficiency produce βs that would be interpreted as large effects

in social science conventions ($|\beta| \gtrsim 0.50$), meaning a one-SD increase in these variables is associated with roughly half to nearly one SD increase in RDC, holding other variables constant. Self-Reported Phishing-Cue Knowledge contributes moderate, positive coefficients across multi-item frames, confirming that explicit cue schemas carry incremental explanatory power beyond general comfort/proficiency. Operational fluency with email clients and general web problem-solving proficiency account for substantial variance in detection confidence. Explicit knowledge of phishing indicators, such as sender/address alignment, link target versus display text discrepancies, and attachment expectations, provides additional, reliable explanatory power beyond general technical comfort.

Cue-checking frequency exhibits small, frame-dependent standardized effects ($\beta \approx 0.20$ - 0.25 when significant). Although routine inspection contributes incrementally to confidence, its magnitude is substantially lower than that of skill- and knowledge-based predictors. Sanctioned confirmation behavior demonstrates limited independent association with retrospective detection confidence once core capability variables are included, reaching statistical significance in only the stimulus-specific frame. These findings suggest that confirmation operates primarily as a behavioral safeguard rather than as a determinant of confidence per se.

Personality indicators explain secondary variance. Emotional Stability and Openness show small positive β's in several frames; Extraversion appears once; Agreeableness/Conscientiousness hover near zero. In variance terms, personality-only models yield low R^2 (≈ 0.08 - 0.20), far less than blocks that include skill/knowledge. Personality traits exert modest and context-dependent effects on detection confidence; however, capability-based predictors account for substantially greater variance in reported confidence levels. Finally, tenure proxies, years of email/Internet/work, are negligible or inconsistent. Statistically, their β's are near zero and frequently non-significant; practically, counting years is a poor proxy for the capabilities that drive confidence.

These patterns hold under robustness checks. Multicollinearity is modest (VIFs < 5), indicating that Email Comfort and Internet Proficiency, though conceptually related, are empirically separable; heteroskedasticity-robust (HC3) errors do not alter point estimates; influence diagnostics show no leverage points that change inference; alternative codings of confirmation and the addition of contextual covariates leave signs and magnitudes essentially unchanged. The relative ordering of predictors remains stable across alternative model specifications, robustness checks, and diagnostic adjustments, indicating that the observed effects are not artifacts of estimation procedures.

Programmatically, the implications are direct. First, train to fluency: brief, hands-on repetitions in the actual mail client and browser, hover, expand headers, compare sender identity to address, and evaluate link targets, shift the constructs that mathematically move confidence the most. Second, rehearse explicit cue schemas so they are retrieved under load; treating "cue knowledge" as a practiced checklist aligns with the moderate β's observed for Self-Reported Phishing-Cue

Knowledge. Third, instrument low-friction confirmation for when cues conflict, one-click ticketing, no-reply banners, directory lookups, but recognize that this chiefly improves behavioral safety rather than raw confidence. Fourth, de-emphasize generic training volume and years-based metrics as KPIs for confidence; the math shows they do not predict RDC once fluency and cue schemas are accounted for. Finally, monitor calibration: track time-to-verification and channel choice alongside accuracy to identify pockets of high confidence/low discrimination, a risk pattern consistent with the literature and implied by our ordering effect.

6. Limitations and Future Work

This study has several limitations that bound inference and generalizability. First, the sample comprises a single-site, professional cohort of DoD engineers (N = 97) responding to an authority-framed stimulus; both the organizational context and persuasion principle may constrain external validity to other sectors, roles, or attack framings. Second, several constructs relied on single-item indicators for field feasibility, which caps reliability and likely attenuates true effects; although internal consistency for the RDC composite was acceptable ($\alpha = 0.753$), measurement error in predictors probably biases coefficients toward zero, a pattern consistent with established psychometric and attenuation principles in scale development research [41] [42]. Third, the design is cross-sectional and retrospective, introducing potential recall bias, social desirability, and common-method variance; despite robustness checks, causal direction cannot be asserted. Fourth, mandatory security training and engineering workload may induce range restriction (e.g., uniformly high fluency), suppressing variance in key predictors and limiting detection of moderation by workload or role. Fifth, nonresponse and survivorship effects are possible (e.g., security-engaged employees may have been more likely to participate), and demographic covariates were unavailable due to HR constraints, limiting subgroup analyses. Sixth, although the study models predictors of retrospective detection confidence (RDC), it does not directly link RDC to observed behavioral outcomes within the same dataset. While process indicators were collected, the present analysis focuses on confidence as a proximal belief state rather than on click, reply, or verification outcomes. As such, conclusions pertain to determinants of confidence rather than demonstrated behavioral protection or calibration.

Future work should directly link confidence to behavior, pair RDC with telemetry from phishing simulations (click/reply outcomes, time-to-verification, channel choice), and quantify calibration using over/under-confidence indices, Brier scores, and signal-detection measures (d′, c) to identify when confidence is protective versus risky. Causal tests should isolate the top levers via randomized interventions that independently manipulate email-fluency drills, cue-schema rehearsal, and confirmation UX (such as no-reply to banners and one-click ticketing), estimating pre/post changes in standardized coefficients and model fit, with heterogeneity-of-treatment-effect analyses by role and workload. Generalization

should be evaluated through multi-site, multi-industry replications that rotate persuasion principles, including urgency/scarcity, reciprocity, and social proof, to test the stability of the observed effect ordering. Measurement should move from single items to short, validated multi-item scales, add prospective confidence alongside RDC, and include demographics to enable moderation tests across experience bands and job families. Process modeling should examine capability/knowledge-to-RDC-to-verification/action pathways using SEM or sequential regression with sensitivity checks for omitted-variable bias. Finally, policy-oriented studies should translate a one–standard-deviation gain in email comfort or cue knowledge into expected reductions in high-risk actions to inform budgeting and governance, while preregistration, de-identified codebooks, and analysis sharing advance transparency and address privacy and fairness concerns when instrumenting inbox workflows at scale.

7. Conclusions

This study sets out to explain what drives users' retrospective confidence in phishing detection and, specifically, whether confidence is better explained by who people are, what they know and can do, or how they work through cues in the moment. The results consistently show that confidence is primarily a function of trainable capability and knowledge, not tenure or generic exposure. In standardized models, email fluency (comfort using email), perceived Internet problem-solving proficiency, and explicit Self-Reported Phishing-Cue Knowledge carry the largest coefficients (typically $\beta \approx 0.50$ - 0.90 across frames), while cue-checking frequency adds a small, frame-specific increment, and personality contributes secondary variance. Years of email/Internet/work experience and recent training volume do not reliably explain additional confidence once those core levers are included.

Detection confidence is highest among individuals demonstrating fluency with email tools, proficiency in Internet problem-solving, and explicit knowledge of diagnostic phishing cues. Habitual cue-checking contributes marginally, and personality traits exert smaller, context-dependent effects. Years of experience and generic training volume do not reliably predict confidence once core skill- and knowledge-based variables are included in the model.

The answer to the primary research question is that retrospective detection confidence is driven primarily by trainable capability variables: email fluency, perceived Internet proficiency, and explicit knowledge of diagnostic phishing cues. Cue-checking behavior contributes a smaller, frame-dependent increment, and personality factors account for secondary variance. Years-based tenure and generic training volume do not reliably explain additional confidence once core skill and knowledge variables are included.

The answer to Subsidiary Research Question 1 (user characteristics): Among user characteristics, perceived capability matters most. Comfort with email and Internet proficiency show strong, positive effects; certain traits (emotional stabil-

ity and openness) contribute to smaller, context-dependent boosts; extraversion appears in one frame only; agreeableness and conscientiousness do not matter. Years of experience (email, Internet, work) are negligible or inconsistent predictors of retrospective detection confidence once capability-based variables are included. Overall, the findings suggest that skill fluency and explicit cue knowledge are more reliable determinants of confidence than tenure or generic training exposure.

The answer to Subsidiary Research Question 2 (cues) is that explicit knowledge of diagnostic phishing indicators is a consistent and meaningful predictor of retrospective detection confidence. Routine cue-checking behavior contributes to a smaller, frame-dependent increment. In contrast, use of a sanctioned confirmation channel, while behaviorally protective, does not independently generate reported confidence once core skill and knowledge variables are included in the model. Only the item capturing whether a consulted opinion substantively influenced the decision reached significance, and only in the stimulus-specific frame.

From an organizational perspective, these findings suggest prioritizing capability-based interventions over tenure-based metrics. Interventions emphasizing hands-on fluency under realistic workload conditions, explicit cue-schema rehearsal to support automatic retrieval, and friction-reduced out-of-thread verification align with the strongest predictors identified in this study. Evaluation frameworks may benefit from incorporating calibration metrics, such as time-to-verification and channel choice, alongside accuracy to distinguish justified from miscalibrated confidence. Emphasis on demonstrated capability and calibrated confidence may provide more actionable indicators than measures based solely on training volume or years of experience. Under such conditions, detection confidence may function as a protective asset rather than a source of risk within operational email environments.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Verizon (2025) 2025 Data Breach Investigations Report (DBIR). Verizon. https://www.verizon.com/business/resources/reports/dbir/

[2] Anti-Phishing Working Group (APWG) (2025) Phishing Activity Trends Report, 4th Quarter 2024. https://docs.apwg.org/reports/apwg_trends_report_q4_2024.pdf

[3] Dhamija, R., Tygar, J.D. and Hearst, M. (2006). Why Phishing Works. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Montréa, 22-27 April 2006, 581-590. https://doi.org/10.1145/1124772.1124861

[4] Jagatic, T.N., Johnson, N.A., Jakobsson, M. and Menczer, F. (2007) Social Phishing. *Communications of the ACM*, **50**, 94-100. https://doi.org/10.1145/1290958.1290968

[5] Wright, R.T. and Marett, K. (2010) The Influence of Experiential and Dispositional Factors in Phishing: An Empirical Investigation of the Deceived. *Journal of Management Information Systems*, **27**, 273-303.

https://doi.org/10.2753/mis0742-1222270111

[6] Bandura, A., Freeman, W.H. and Lightsey, R. (1999) Self-Efficacy: The Exercise of Control. *Journal of Cognitive Psychotherapy*, **13**, 158-166. https://doi.org/10.1891/0889-8391.13.2.158

[7] Rhee, H., Kim, C. and Ryu, Y.U. (2009) Self-efficacy in Information Security: Its Influence on End Users' Information Security Practice Behavior. *Computers & Security*, **28**, 816-826. https://doi.org/10.1016/j.cose.2009.05.008

[8] Parsons, K., McCormac, A., Butavicius, M., Pattinson, M. and Jerram, C. (2014) Determining Employee Awareness Using the Human Aspects of Information Security Questionnaire (HAIS-Q). *Computers & Security*, **42**, 165-176. https://doi.org/10.1016/j.cose.2013.12.003

[9] Brickley, J.C. (2022) What Drives User Retrospective Confidence Levels Towards Phishing Detection Behavior. Ph.D. Thesis, New Jersey City University.

[10] Souppaya, K. and Scarfone, K. (2013) Guide to Malware Incident Prevention and Handling for Desktops and Laptops. NIST Special Publication 800-83 Rev. 1, National Institute of Standards and Technology.

[11] Cichonski, P., Millar, T., Grance, T. and Scarfone, K. (2012) Computer Security Incident Handling Guide. NIST Special Publication 800-61 Rev. 2, National Institute of Standards and Technology.

[12] Li, N. and Wu, D. (2005) A Framework for Secure Electronic Commerce Transactions. *Decision Support Systems*, **39**, 271-283.

[13] Alseadoon, I. (2014) Phishing Detection Behavior: A User-Centered Model. Ph.D. Thesis, University of New South Wales.

[14] Wu, M., Miller, R.C. and Garfinkel, S.L. (2006) Do Security Toolbars Actually Prevent Phishing Attacks? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Montréal, 22-27 April 2006, 601-610. https://doi.org/10.1145/1124772.1124863

[15] Pandove, G., Maity, R. and Sharma, S. (2010) E-Mail Spoofing: Techniques and Countermeasures. *Proceedings of ICCCT*, Guangxi, 13-14 October 2010, 394-398.

[16] Akbar, M. (2014) Phishing E-Mails and Cialdini's Principles of Persuasion. Master's Thesis, University of Twente.

[17] Cialdini, R.B. (2009) Influence: Science and Practice. 5th Edition, Pearson.

[18] Jansen, J. and Leukfeldt, R.E. (2016) Phishing and Malware Attacks on Online Banking Customers in the Netherlands: A Qualitative Analysis of Factors Leading to Victimization. *International Journal of Cyber Criminology*, **10**, 79-91.

[19] Workman, M. (2008) A Test of Intervention for Security Threats: Decision Biases in Computer Security. *Decision Sciences*, **39**, 615-643.

[20] Guéguen, N., Pascual, A. and Dagot, J.L. (2010) Similarity and Compliance to a Request: A Field Study on Pedestrians. *Social Behavior and Personality*, **38**, 453-458.

[21] Guadagno, R.E., Muscanell, N.L., Rice, L.M. and Roberts, N. (2013) Social Influence Online: The Impact of Social Validation and Likability on Compliance. *Psychology of Popular Media Culture*, **2**, 51-60. https://doi.org/10.1037/a0030592

[22] Lawson, P., Pearson, C.J., Crowson, A. and Mayhorn, C.B. (2020) Email Phishing and Signal Detection: How Persuasion Principles and Personality Influence Response Patterns and Accuracy. *Applied Ergonomics*, **86**, Article ID: 103084. https://doi.org/10.1016/j.apergo.2020.103084

[23] Xun, D., Clark, J.A. and Jacob, J. (2008) Modelling User-Phishing Interaction. 2008

Conference on Human System Interaction, Krakow, 25-27 May 2008, 627-632. https://doi.org/10.1109/HSI.2008.4581513

[24] Aburrous, M., Hossain, M.A., Dahal, K. and Thabtah, F. (2010) Intelligent Phishing Detection System for E-Banking Using Fuzzy Data Mining. *Expert Systems with Applications*, **37**, 7913-7921. https://doi.org/10.1016/j.eswa.2010.04.044

[25] Vishwanath, A., Herath, T., Chen, R., Wang, J. and Rao, H.R. (2011) Why Do People Get Phished? Testing Individual Differences in Phishing Vulnerability within an Integrated, Information Processing Model. *Decision Support Systems*, **51**, 576-586. https://doi.org/10.1016/j.dss.2011.03.002

[26] Petty, R.E. and Cacioppo, J.T. (1986) Communication and Persuasion: Central and Peripheral Routes to Attitude Change. Springer.

[27] Grazioli, S. (2004) Where Did They Go Wrong? An Analysis of the Failure of Knowledgeable Internet Consumers to Detect Deception over the Internet. *MIS Quarterly*, **28**, 387-422.

[28] Higgins, E.T. (1996) Knowledge Activation: Accessibility, Applicability, and Salience. In: Higgins, E.T. and Kruglanski, A.W., Eds., *Social Psychology*: *Handbook of Basic Principles*, Guilford, 133-168.

[29] Rietbergen, A. (2015) An Eye-Tracking Study on What Receives Attention in E-Mails in an Intercultural Context. Master's Thesis, Radboud University.

[30] Jain, A.K. and Gupta, B.B. (2017) Phishing Detection: Analysis of Visual Similarity Based Approaches. *Security and Communication Networks*, **2017**, Article ID: 5421046. https://doi.org/10.1155/2017/5421046

[31] Ho, S.Y. and Bodoff, D. (2014) The Effects of Web Personalization on User Attitude and Behavior: An Integration of the Elaboration Likelihood Model and Consumer Search Theory. *MIS Quarterly*, **38**, 497-520. https://doi.org/10.25300/misq/2014/38.2.08

[32] Wang, J., Li, Y. and Rao, H.R. (2016) Overconfidence in Phishing Email Detection. *Journal of the Association for Information Systems*, **17**, 759-783. https://doi.org/10.17705/1jais.00442

[33] Fazio, R.H. and Zanna, M.P. (1978) On the Predictive Validity of Attitudes: The Roles of Direct Experience and Confidence. *Journal of Personality*, **46**, 228-243. https://doi.org/10.1111/j.1467-6494.1978.tb00177.x

[34] Moore, D.A. and Healy, P.J. (2008) The Trouble with Overconfidence. *Psychological Review*, **115**, 502-517. https://doi.org/10.1037/0033-295x.115.2.502

[35] Berger, I.E. (1992) The Nature of Attitude Accessibility and Attitude Confidence: A Triangulated Experiment. *Journal of Consumer Psychology*, **1**, 103-123. https://doi.org/10.1016/s1057-7408(08)80052-6

[36] Berger, I.E. and Mitchell, A.A. (1989) The Effect of Advertising on Attitude Accessibility, Attitude Confidence, and the Attitude-Behavior Relationship. *Journal of Consumer Research*, **16**, 269-279. https://doi.org/10.1086/209213

[37] Busey, T.A., Tunnicliff, J., Loftus, G.R. and Loftus, E.F. (2000) Accounts of the Confidence-Accuracy Relation in Recognition Memory. *Psychonomic Bulletin & Review*, **7**, 26-48. https://doi.org/10.3758/bf03210724

[38] Stone, D.N. (1994) Overconfidence in Initial Self-Efficacy Judgments: Effects on Decision Processes and Performance. *Organizational Behavior and Human Decision Processes*, **59**, 452-474. https://doi.org/10.1006/obhd.1994.1069

[39] Moores, T.T. and Chang, J.C. (2009) Self-Efficacy, Overconfidence, and the Negative

Effect on Subsequent Performance: A Field Study. *Information & Management*, **46**, 69-76. https://doi.org/10.1016/j.im.2008.11.006

[40] Cramer, R.J., Neal, T.M.S. and Brodsky, S.L. (2009) Self-Efficacy and Confidence: Theoretical Distinctions and Implications for Trial Consultation. *Consulting Psychology Journal: Practice and Research*, **61**, 319-334. https://doi.org/10.1037/a0017310

[41] Alseadoon, I., Othman, M.H., Chan, A. and Foo, S. (2017) Typology of Phishing Victims Based on Personality Traits and Other Factors. *Computers in Human Behavior*, **66**, 1-13.

[42] Srivastava, S., John, O.P., Gosling, S.D. and Potter, J. (2003) Development of Personality in Early and Middle Adulthood: Set Like Plaster or Persistent Change? *Journal of Personality and Social Psychology*, **84**, 1041-1053. https://doi.org/10.1037/0022-3514.84.5.1041