Computers in Human Behavior 186 (2027) 109156

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

At the point of risk: Can AI-assisted decision support reduce digital fraud vulnerability across age groups and scam types?☆,☆☆

Yicheng Sun *, Jacky Keung, Hi Kuen Yu *, Yuchen Cao

Department of Computer Science, City University of Hong Kong, 999077, Hong Kong, China

ARTICLE INFO

Keywords:
Digital fraud
Human–AI interaction
AI-assisted decision support
Fraud susceptibility
Risky behavioral intention
Age differences

ABSTRACT

Digital communication has broadened fraud's reach and complexity, making detection more urgent and cognitively demanding. Yet most studies examine static risk profiles or single scams, leaving real-time support in realistic settings underexplored. To address this gap, this study examined an AI-assisted decision support framework for digital fraud prevention. Using a 2 × 2 online lab-in-the-field experimental design, we compared an AI-assisted condition with a control condition across a younger group and an older group. A total of 72 participants completed a 10-day study involving ecologically realistic simulated fraud scenarios across six scam categories: financial/payment fraud, cyber/technical fraud, social relationship scams, consumer fraud, government or institutional impersonation fraud, and social media scams. The AI system provided real-time risk assessments, explanations of suspicious cues, and recommended actions. The results showed that AI-assisted support improved fraud detection accuracy and reduced risky behavioral intention. Younger and older groups also showed different scam-specific vulnerability patterns, and AI support reduced participants' willingness to engage in risky actions in response to fraudulent communications in the scam categories to which each group was most vulnerable. Post-study questionnaire ratings and semi-structured interviews further indicated that participants generally perceived the system as useful and realistic, while trust depended on whether the AI output was interpretable and actionable. These findings demonstrate the potential of AI-assisted decision support for digital fraud prevention and highlight the importance of interpretable and actionable support that accounts for age-specific vulnerability patterns, offering practical guidance for the design of digital safety interventions in everyday online environments.

1. Introduction

The rapid integration of digital technologies into financial, social, and institutional life has transformed the ways in which individuals communicate, transact, and make everyday decisions. At the same time, this transformation has expanded both the reach and the complexity of fraud (Jair & Mustafa, 2025; Reshetnikova et al., 2021). In China alone, police solved approximately 258,000 telecom and online fraud cases in 2025, while more than 21.7 million RMB in fraud-related funds were frozen through emergency measures, underscoring both the scale and urgency of the problem (State Council Information Office of the People's Republic of China, 2026). Contemporary scams no longer rely solely on isolated deception attempts; instead, they increasingly appear through emails, text messages, social media platforms, app notifications, and other digitally mediated channels (Baki & Verma, 2022; Bele, 2020), often combining technical manipulation with urgency, authority cues, emotional pressure, and requests for immediate action (Dupuis et al., 2023; Fan & Yu, 2022). As a result, fraud-related decision making has become an increasingly time-sensitive and cognitively demanding activity for ordinary users navigating everyday digital environments.

Recent research suggests that fraud vulnerability cannot be explained by any single factor alone. Rather, susceptibility to fraud emerges from the interaction among characteristics of the fraud itself, individual-level differences, and broader situational conditions (Bar Lev et al., 2022). The recent systematic review by Dadà et al. (2025) noted that fraud victims cannot be treated as a homogeneous group and that victimization appears to result from the interplay of scam type, individual characteristics, and contextual conditions, with online environments now playing a particularly important role in ex-

☆ This article is part of a Special issue entitled: 'DFLS' published in Computers in Human Behavior.

☆☆ This work is supported in part by the General Research Fund of the Research Grants Council of Hong Kong and the Research Funds of the City University of Hong Kong (6000796, 9229109, 9229098, 9220103, 9229029).

* Corresponding authors.
E-mail addresses: yicsun2-c@my.cityu.edu.hk (Y. Sun), Jacky.Keung@cityu.edu.hk (J. Keung), hikuenyu2-c@my.cityu.edu.hk (H.K. Yu), cynthcao2-c@my.cityu.edu.hk (Y. Cao).

https://doi.org/10.1016/j.chb.2026.109156
Received 29 March 2026; Received in revised form 11 August 2026; Accepted 16 August 2026
Available online 28 August 2026

posure and vulnerability. The same review further showed that the most consistently supported risk factors across studies include low conscientiousness, higher impulsivity, reduced cognitive functioning, and social isolation, while also emphasizing that people of all ages may be vulnerable depending on the form and context of the scam (Burton et al., 2022; Colautti et al., 2025; Dadà et al., 2025; DeLiema et al., 2023). These findings highlight the need to understand fraud vulnerability not only in terms of individual-level characteristics, but also as a dynamic decision-making challenge shaped by person-level, task-level, and environmental factors.

Existing fraud-prevention research also provides important evidence that warning and awareness-based interventions can reduce susceptibility under some conditions, but also shows that their effects depend strongly on timing, delivery format, target population, and decision context. Experimental evidence suggests that forewarning can reduce fraud susceptibility among vulnerable consumers, particularly when warnings help individuals recognize risk before engaging with a fraudulent offer (Scheibe et al., 2014). At the same time, emotional arousal can increase susceptibility to fraud among both younger and older adults, indicating that prevention efforts must address not only knowledge deficits but also the affective and situational pressures under which fraudulent decisions are made (Kircanski et al., 2018). More recent work has also questioned the effectiveness of broad awareness campaigns: Jensen et al. found little evidence that a mass-message financial fraud awareness campaign significantly reduced fraud victimization, suggesting that passive or one-off messaging may be insufficient (Jensen et al., 2025). Research on fraud-prevention advice for older adults further indicates that effective delivery may require trusted, audience-sensitive, and multi-channel communication rather than relying only on websites or generic public messages (Button et al., 2024). Together, these studies suggest that fraud prevention is most likely to be effective when support is timely, specific, trusted, and available at the point of decision.

Despite this growing body of work, several important limitations remain. First, much of the existing literature has focused on identifying relatively static predictors of victimization, such as demographic characteristics, psychological traits, or prior experiences, rather than examining how individuals make fraud-related decisions in real time (James et al., 2014; Junger et al., 2023; Norris et al., 2019). Second, many studies have centered on specific populations, age groups, or particular scam types, often relying on context-bound or retrospective victimization data, which makes it difficult to understand how people respond across diverse digital fraud contexts (Irvin-Erickson, 2024; Shang et al., 2022; Ueno et al., 2022). Third, as noted in the review by Dadà et al. (2025), experimental designs in this area remain relatively limited, and there is a continuing need for more standardized, digitally aware, and context-sensitive approaches to the study of fraud vulnerability. In other words, current research has been more successful in describing who may be vulnerable than in testing how such vulnerability might be reduced at the moment when risk is encountered.

This limitation is especially consequential because fraud exposure is often highly time-sensitive. Fraudulent messages are frequently constructed around deceptive and persuasive cues, including urgency, authority, emotional pressure, reward promises, suspicious links, and requests for credentials or immediate action (Butavicius et al., 2022; Dadà et al., 2025; Ferreira et al., 2015; Ferreira & Teles, 2019). These cues can pressure targets into making hasty decisions before they have sufficient time to verify suspicious claims, and prior work has shown that reduced response time can impair the ability to distinguish legitimate from fraudulent communications (Butavicius et al., 2022; Dadà et al., 2025; Jones et al., 2019). In practice, individuals may attempt to seek advice from family members, friends, or other trusted contacts when confronted with uncertain or potentially fraudulent messages, yet such support is not always immediately available. Existing research suggests that limited opportunities to verify information or consult trusted others can increase vulnerability, particularly in contexts marked by social isolation or reduced support networks (Du & Chen, 2023; Kadoya et al., 2021; Xiang et al., 2020). In this respect, recent advances in artificial intelligence create a promising new possibility: when human advice or verification support is unavailable, an AI-based decision support system can provide on-demand assistance at the precise moment of exposure by identifying suspicious cues, estimating fraud risk, and recommending safer responses. This makes AI a potentially valuable intervention resource in real-world digital fraud settings.

Against this background, the present study examines whether an AI-assisted decision support framework can reduce susceptibility to digital fraud in ecologically valid settings. Rather than treating fraud vulnerability only as a matter of individual-level characteristics, we conceptualize fraud detection as a situated decision-making process shaped by user characteristics, deceptive task features, and contextual support. This perspective is especially relevant to digital fraud because suspicious communications are often encountered under time pressure, uncertainty, and limited opportunities for independent verification. A 2×2 online lab-in-the-field experimental design was chosen because it allowed us to examine both the effect of AI-assisted support and age-related differences within the same design. The comparison between AI-assisted and control conditions provided a direct test of whether real-time decision support improved fraud-related judgments beyond unaided human judgment, while the comparison between younger and older groups allowed us to examine whether intervention effects differed across age groups. The naturalistic and repeated-exposure design was also important because fraud-related decisions often occur in everyday communication environments rather than in isolated laboratory tasks.

In a 10-day online lab-in-the-field experiment conducted in mainland China and Hong Kong, participants encountered simulated but ecologically realistic fraud scenarios spanning six fraud categories: financial/payment, cyber/technical, social-relationship, consumer, government/institutional impersonation, and social-media fraud. In this study, financial/payment fraud refers to attempts to obtain money or financial credentials; cyber/technical fraud refers to phishing-like security alerts, fake login requests, or technical-support messages designed to obtain credentials or personal information; social-relationship fraud refers to messages exploiting interpersonal trust or emotional pressure; consumer fraud refers to fraudulent shopping, delivery, refund, or promotional messages; government/institutional impersonation fraud refers to messages falsely claiming official authority; and social-media fraud refers to fraudulent private messages or platform-based messages that redirect users to external sites or solicit sensitive information. Detailed operational definitions and examples are provided in Table 4. By combining repeated behavioral measures with post-study questionnaire ratings and semi-structured interviews, the study moves beyond static risk profiling toward a more applied, context-sensitive, and intervention-oriented investigation of fraud-related decision making. The mixed-methods component was included to complement behavioral outcomes with participants' subjective evaluations and interview accounts, thereby helping to explain how participants interpreted, trusted, and used the AI support system.

Drawing on the Person–Task Fit framework, as developed in the next section, the study focuses on three linked issues: whether AI-assisted decision support improves fraud detection and reduces risky behavioral intention, whether younger and older groups show different vulnerability patterns across fraud categories, and how users evaluate the usefulness, trustworthiness, realism, and confidence-related effects of such support after repeated exposure. These issues are expressed in the following research questions:

- **RQ1:** Does access to an AI-assisted decision support system improve fraud detection accuracy and reduce risky behavioral intention compared with unaided human judgment?

Table 1
Mapping between person–task fit components and integrated theoretical perspectives.

PTF framework component	Integrated theory or perspective	Role in the present study
Person factors	Individual-difference perspective on fraud vulnerability; age-related digital literacy and risk-perception research	Explains how users' age, digital habits, prior fraud exposure, confidence, and perceived risk may shape their ability to identify suspicious communications and resist unsafe actions.
Task factors	Deceptive-cue processing perspective; scam persuasion and emotional-arousal research	Explains how fraud messages create task demands through urgency cues, authority impersonation, emotional pressure, reward promises, suspicious links, and low-cost requests that may increase compliance.
Contextual support	AI-assisted decision-support perspective; just-in-time warning and explainable intervention research	Explains how real-time AI support may improve person–task fit by providing risk assessments, explanations of suspicious cues, and recommended actions at the moment of decision.
Overall person–task fit	Person–Task Fit framework	Provides the overarching logic for understanding fraud vulnerability as an interaction between user characteristics, deceptive task demands, and available contextual support.

- **RQ2:** How do fraud vulnerability patterns differ between the younger and older groups across fraud categories, and can AI-assisted decision support reduce susceptibility to the fraud categories to which each age group is most vulnerable?
- **RQ3:** How do participants perceive and experience the AI-assisted decision support system in digital fraud detection, in terms of usefulness, trust, realism, and confidence?

2. Theoretical framework: Person–Task Fit and AI-assisted fraud detection

The present study is guided by the Person–Task Fit (PTF) framework, which proposes that decision quality depends on the alignment between individual characteristics and task demands within a specific context (Wronska et al., 2019). This framework is particularly relevant to digital fraud because suspicious communications are rarely evaluated under ideal conditions. Instead, users often need to judge messages under uncertainty, time pressure, incomplete information, and deliberate manipulation (Ferreira et al., 2015; Ferreira & Teles, 2019). In such settings, vulnerability should not be understood only as a stable attribute of the individual. It can also emerge when the demands of a deceptive communication exceed the user's immediate ability or available resources for evaluating it.

Applied to digital fraud detection, the PTF framework highlights three interrelated elements: person factors, task factors, and contextual support. Person factors include characteristics such as age, digital literacy, prior fraud exposure, confidence, risk perception, and cognitive resources. Previous research suggests that these factors can shape fraud vulnerability, although age-related effects are not always uniform across contexts (Baki & Verma, 2022; Dadà et al., 2025; James et al., 2014; Kemp & Erades Pérez, 2023; Shang et al., 2022). For example, studies on fraud and phishing susceptibility have shown that older group may face particular risks when scams involve authority, trust, cognitive burden, or lower digital literacy, while younger group may also be vulnerable in highly familiar digital environments such as social media or online consumer platforms. This suggests that age should not be treated as a simple indicator of greater or lower vulnerability, but as a factor that may interact with the type of deceptive task.

Task factors refer to the features of the fraudulent communication itself. These include urgency cues, authority impersonation, emotional pressure, reward promises, suspicious links, requests for credentials, and demands for small but immediate payments. Such cues are important because they increase the cognitive and emotional demands of fraud detection. This is consistent with experimental evidence showing that emotional arousal can increase susceptibility to fraud among both younger and older (Kircanski et al., 2018). A message may appear routine, low-cost, or familiar while still directing the user toward unsafe behavior. From a PTF perspective, fraud vulnerability is therefore more likely when the deceptive cues embedded in a message match the recipient's expectations or routines and are not recognized as warning signs.

Contextual support refers to the resources available to the user at the moment of decision. In everyday settings, users may try to verify suspicious messages by consulting family members, friends, official websites, or customer service channels. However, such support is not always immediately available. AI-assisted decision support can therefore be conceptualized as a form of contextual support that helps improve person–task fit. Rather than replacing human judgment, the system provides real-time risk assessments, explanations of suspicious cues, and recommended actions. These forms of support may help users identify task-relevant warning signs, reduce uncertainty, and choose safer responses when they encounter potentially fraudulent communications.

The PTF framework also explains why AI-assisted support may not have identical effects across all users or fraud categories. If younger and older groups differ in their digital habits, trust cues, prior exposure, or response strategies, then the same fraud category may create different forms of person–task mismatch for each group. Similarly, AI support may be especially useful when it makes hidden or underestimated risk cues more visible, such as suspicious links, false authority claims, emotional manipulation, or urgency pressure. To make the theoretical integration more explicit, Table 1 summarizes how the three components of the Person–Task Fit framework are connected with the theoretical and empirical perspectives used in this study. This mapping clarifies that PTF is used as the overarching framework, while specific theories and prior research inform how person-related vulnerability, scam-task characteristics, and AI-assisted contextual support are operationalized in the present study.

On this basis, the present study uses PTF as an analytical framework to examine whether AI-assisted decision support improves fraud detection, reduces risky behavioral intention, and mitigates age-specific vulnerability patterns across different fraud categories.

3. Methodology

3.1. Research design

The study adopted a 2 × 2 online lab-in-the-field experimental design to examine the effects of AI-assisted decision support and age group on digital fraud-related decision making. This design combines experimental control with participation under naturalistic digital-use conditions. Experimental settings may vary along several dimensions, including the characteristics of participants, the nature of the task,

the information available to participants, the stakes involved, and the environment in which decisions are made (Harrison & List, 2004). Accordingly, experiments conducted outside conventional laboratory settings while retaining researcher-controlled tasks have been described as framed field, extra-laboratory, or lab-in-the-field experiments (Charness et al., 2013; Gangadharan et al., 2022). Online environments can also provide a suitable setting for such experimental designs by allowing controlled interventions to be evaluated under conditions that more closely reflect participants' everyday digital activities (Chen & Konstan, 2015).

In the present study, the fraud scenarios and AI-assisted intervention were systematically designed and delivered through a study-specific web platform to ensure consistency across experimental conditions. At the same time, participants completed the study over a 10-day period using their own smartphones or computers in their usual environments, rather than attending a single session in a physical laboratory. Fraud scenarios were made available at irregular intervals and presented through communication formats resembling commonly encountered emails, text messages, app notifications, and chat-style messages. This combination of controlled stimulus presentation and repeated participation in participants' everyday digital environments enabled the study to maintain internal experimental control while increasing the contextual realism of fraud-related decision making.

3.2. Participants

The study was conducted in China, with participants recruited from both mainland China and Hong Kong. A total of 72 participants were enrolled in the study. To examine potential age-related differences in responses to AI-assisted fraud detection, participants were stratified into two age groups: **younger group** aged 18–30 years and **older group** aged 50 years and above. The final sample comprised 36 younger participants and 36 older participants. Participants were recruited through university mailing lists, community organizations, and social media advertisements. These recruitment channels were used to reach individuals with different everyday digital communication habits and to ensure that both age groups included participants with regular exposure to common digital platforms. Because the study was designed as a controlled experiment rather than a population survey, recruitment prioritized balanced assignment across age group and decision-support condition, as well as participants' ability to complete repeated fraud-detection tasks over the 10-day study period. The resulting sample therefore provides a basis for examining experimental differences between the study groups, while broader population-level generalization should be made with caution.

Within each age group, participants were assigned to either the AI-assisted condition or the control condition, resulting in four equally sized groups: younger group without AI support ($n = 18$), younger group with AI support ($n = 18$), older group without AI support ($n = 18$), and older group with AI support ($n = 18$). Prior to group assignment, all participants completed a brief pre-assessment interview and orientation session designed to assess their basic familiarity with everyday digital communication tools and to confirm that they were able to engage with the study tasks. This preliminary step also allowed the research team to check for major discrepancies in baseline digital usage and communication habits, so that the two experimental conditions within each age group were as comparable as possible before assignment. This allocation procedure helped maintain balance across age and intervention conditions and supported subsequent comparisons of both main and interaction effects.

Eligibility criteria were established to ensure that participants could meaningfully engage with the study tasks in a realistic digital environment. Specifically, participants were required to (a) use a smartphone or computer on a regular basis, (b) be able to receive and respond to digital messages during the 10-day study period, and (c) have no prior participation in similar fraud-detection experiments that might bias their responses. Because the study focused on fraud-related decision-making in ecologically valid contexts, all participants were expected to possess sufficient familiarity with common digital communication channels such as text messaging, email, and social media.

Table 2 summarizes the demographic and background characteristics of the sample, including age, gender, educational attainment, employment status, occupational category, living arrangement, primary device use, daily digital usage, and prior exposure to suspicious or potentially fraudulent messages. In detail, the two age groups differed primarily in age-related and employment-related characteristics, while both groups demonstrated regular engagement with digital communication technologies. In addition, a majority of participants in both groups reported prior exposure to suspicious or potentially fraudulent messages, suggesting that the sample possessed sufficient familiarity with everyday fraud-related communication contexts.

Participants also completed a baseline questionnaire prior to the experimental exposure phase. This assessment served two purposes: first, to characterize participants' initial levels of fraud-relevant capabilities and experiences, and second, to examine whether the four groups were broadly comparable before the intervention began. The questionnaire consisted of structured quantitative items, including categorical background questions, yes/no items, multiple-choice items, and five-point Likert-scale items. It assessed digital literacy, prior fraud exposure, previous scam-related harm, risk perception, confidence in detecting suspicious messages, and everyday digital communication habits. No open-ended qualitative questions were included in the baseline questionnaire; qualitative data were collected separately through the post-study semi-structured interviews. A copy of the baseline questionnaire is provided in Appendix.

First, *digital literacy* was evaluated using self-report items concerning participants' familiarity with common digital communication tools, their ability to identify suspicious links or messages, and their confidence in verifying the authenticity of online content. Second, *prior fraud exposure* was assessed by asking whether participants had previously encountered phishing emails, suspicious payment requests, impersonation attempts, fraudulent links, or scam-related messages through email, text messaging, phone calls, or social media. Third, *previous scam-related harm* was assessed by asking whether participants had previously been personally affected by a scam, such as through financial loss, disclosure of personal information, clicking a fraudulent link, or taking another action that led to a negative consequence. Fourth, *risk perception* was measured through items examining the extent to which participants viewed digital communication as potentially risky, believed themselves to be vulnerable to online scams, and considered digital fraud to be a realistic threat in everyday life. In addition, the questionnaire collected background information on participants' digital communication habits, including the frequency of smartphone and computer use, average daily time spent online, and preferred communication platforms.

These baseline measures were used both to describe the sample and to assess group comparability prior to the intervention. As shown in Table 3, the four groups were broadly comparable on the main baseline variables relevant to fraud-related decision making. No statistically significant group differences were observed for digital literacy, prior fraud exposure, perceived fraud risk, or self-reported confidence in detecting suspicious messages (all $p > .05$), suggesting that participants entered the study with broadly similar levels of relevant experience and awareness. A significant difference was observed only for average daily digital use, with younger participants reporting more time spent online than older participants. This pattern was expected given age-related differences in everyday digital engagement and did not undermine the overall comparability of the experimental groups.

In practical terms, the baseline results indicate that the four experimental groups were broadly comparable before the intervention. Most baseline variables showed small or very small effect sizes, suggesting limited practical imbalance. The only statistically and practically

Table 2
Participant characteristics by age group.

Characteristic	Younger group	Older group	Total
Sample size, *n*	36	36	72
Gender, *n* (%)			
Male	17 (47.2)	16 (44.4)	33 (45.8)
Female	19 (52.8)	20 (55.6)	39 (54.2)
Age, years			
Mean (SD)	24.08 (3.41)	59.67 (6.82)	41.88 (18.67)
Range	18–30	50–72	18–72
Highest educational attainment, *n* (%)			
High school or below	5 (13.9)	8 (22.2)	13 (18.1)
College or vocational diploma	8 (22.2)	12 (33.3)	20 (27.8)
Bachelor's degree	18 (50.0)	12 (33.3)	30 (41.7)
Postgraduate degree	5 (13.9)	4 (11.1)	9 (12.5)
Employment status, *n* (%)			
Student	21 (58.3)	0 (0.0)	21 (29.2)
Employed full-time	9 (25.0)	18 (50.0)	27 (37.5)
Employed part-time	5 (13.9)	4 (11.1)	9 (12.5)
Retired	0 (0.0)	11 (30.6)	11 (15.3)
Other	1 (2.8)	3 (8.3)	4 (5.6)
Occupational category, *n* (%)			
Student	21 (58.3)	0 (0.0)	21 (29.2)
Professional/managerial/office-based	8 (22.2)	13 (36.1)	21 (29.2)
Service/manual/technical	6 (16.7)	9 (25.0)	15 (20.8)
Retired	0 (0.0)	11 (30.6)	11 (15.3)
Other	1 (2.8)	3 (8.3)	4 (5.6)
Living arrangement, *n* (%)			
Living alone	7 (19.4)	9 (25.0)	16 (22.2)
Living with others	29 (80.6)	27 (75.0)	56 (77.8)
Primary device used in daily life, *n* (%)			
Smartphone mainly	20 (55.6)	14 (38.9)	34 (47.2)
Computer mainly	4 (11.1)	9 (25.0)	13 (18.1)
Both equally	12 (33.3)	13 (36.1)	25 (34.7)
Average daily digital use, *n* (%)			
Less than 2 h	3 (8.3)	11 (30.6)	14 (19.4)
2–4 h	10 (27.8)	15 (41.7)	25 (34.7)
More than 4 h	23 (63.9)	10 (27.8)	33 (45.8)
Prior experience with suspicious or fraudulent messages, *n* (%)	24 (66.7)	19 (52.8)	43 (59.7)

Table 3
Baseline assessment results by experimental group with effect sizes and practical significance.

Measure	YA-control	YA-AI	OA-control	OA-AI	*p*	Effect size	Practical significance
Digital literacy, Mean (SD)	3.92 (0.48)	3.88 (0.51)	3.61 (0.57)	3.66 (0.54)	0.205	$\eta^2 = 0.065$	No statistically significant baseline difference; the effect size was moderate, with descriptively lower digital-literacy scores in the older groups and little difference between AI and control conditions within each age group.
Prior fraud exposure, *n* (%)	12 (66.7)	12 (66.7)	9 (50.0)	10 (55.6)	0.669	$V = 0.147$	Small difference; prior fraud exposure was broadly comparable across experimental groups.
Previously affected by a scam, *n* (%)	4 (22.2)	3 (16.7)	5 (27.8)	4 (22.2)	0.979	$V = 0.094$	Very small difference; previous scam-related harm was broadly balanced across experimental groups.
Risk perception, Mean (SD)	3.57 (0.62)	3.61 (0.58)	3.69 (0.55)	3.72 (0.60)	0.860	$\eta^2 = 0.011$	Very small difference; the groups entered the study with similar levels of perceived fraud risk.
Average daily digital use (hours), Mean (SD)	5.11 (1.63)	4.94 (1.58)	3.41 (1.44)	3.56 (1.49)	< 0.001	$\eta^2 = 0.212$	Statistically and practically meaningful group difference; descriptively, daily digital use was higher in the younger groups than in the older groups.
Confidence in detecting suspicious messages, Mean (SD)	3.44 (0.66)	3.39 (0.61)	3.18 (0.59)	3.22 (0.64)	0.532	$\eta^2 = 0.032$	Small difference; baseline confidence in detecting suspicious messages was broadly comparable across groups.

Note. YA = younger group; OA = older group. For continuous variables, omnibus group differences were evaluated using one-way ANOVA, with effect sizes reported as η^2. For categorical variables, Pearson's χ^2 test was used when expected cell frequencies were adequate, with Cramer's V reported as the association effect size. The Fisher–Freeman–Halton exact test was used to determine the *p*-value for previous scam-related harm because of small expected cell frequencies. Practical significance refers to the extent to which baseline differences may represent meaningful group imbalance before the intervention.

meaningful difference was average daily digital use, with younger participants reporting more time online than older participants, which is consistent with expected age-related differences in everyday digital engagement.

Sample size and sensitivity analysis. Given the fixed sample of 72 participants, we conducted a sensitivity analysis to evaluate the minimum effect size that could be reliably detected by the study design (Faul et al., 2009; Lakens, 2022). The analysis was based conservatively on the 2 × 2 between-participant component of the experimental design, with four groups (n = 18 per group), α = .05, statistical power of .80, and one numerator degree of freedom. Using G*Power 3.1 for a fixed-effects ANOVA sensitivity test for main effects and interactions, the minimum detectable effect size was approximately Cohen's f = 0.335, corresponding to a partial η^2 of approximately .101. Thus, the available sample provided reasonable sensitivity for detecting between-participant effects of this magnitude or larger, whereas smaller effects may have been more difficult to detect reliably. This analysis focuses on the between-participant component because repeated scenario-level observations were nested within participants and therefore could not be treated as independent observations for sample-size justification. The repeated-exposure structure was instead accounted for in the multilevel models described below.

3.3. AI-supported fraud scenarios and experimental materials

The experimental materials were designed to simulate realistic exposure to common forms of digital fraud while allowing systematic evaluation of the AI-assisted decision support framework. To achieve this, the study combined two closely related components: (a) a set of structured fraud scenarios representing major categories of digital scams, and (b) an AI-based support mechanism that provided participants with interpretable decision assistance in the AI-assisted condition. This integrated design ensured that the intervention was embedded within task contexts that closely resembled everyday digital communication environments.

The AI-assisted system evaluated in this study was developed as a prototype decision support framework intended to help users assess whether a message or communication attempt was likely to be fraudulent. Rather than providing only a binary judgment, the system generated three forms of support for each scenario. First, it produced a *fraud risk assessment*, that is, a probability-based indication of the likelihood that the presented message was fraudulent. Second, it provided an *explanation of risk indicators*, highlighting suspicious cues embedded in the message, such as urgency language, authority impersonation, emotional manipulation, reward-based appeals, or requests for sensitive financial or personal information. Third, it generated *recommended actions*, offering practical guidance such as avoiding further interaction, verifying the sender through official channels, or refraining from clicking suspicious links. For illustration, a government impersonation scenario might present participants with a message claiming that an unpaid tax penalty required immediate payment to avoid legal sanctions, accompanied by a hyperlink to an external payment page. In the AI-assisted condition, the system could classify the message as *high risk*, identify the presence of *authority cues*, *threat-based urgency*, and a *non-verifiable payment link* as salient risk indicators, and recommend that the participant refrain from interacting with the message and instead confirm its legitimacy through an official institutional contact channel. In this way, the AI output was designed to support not only fraud detection accuracy, but also participants' understanding of why a given communication should be treated as suspicious. Participants in the AI-assisted condition could consult this support before making their final judgment, whereas participants in the control condition completed the same tasks without AI assistance and relied solely on their own judgment.

To ensure that the intervention remained theoretically grounded and consistent across conditions, the AI outputs were directly aligned with the deceptive cue framework used to construct the fraud scenarios. As shown in Table 5, the experimental materials incorporated a set of recurrent deceptive cues commonly found in digital fraud, including urgency cues, authority cues, reward cues, emotional manipulation, and suspicious hyperlinks. These cues represented the key *task characteristics* in the Person–Task Fit framework, as they shaped the cognitive demands of fraud detection and decision making. The role of the AI system was therefore not merely to classify a message as fraudulent, but to support participants in recognizing and interpreting these task-relevant signals.

Each scenario was classified as fraudulent based on commonly recognized scam indicators in digital fraud research, including requests for financial transfers, solicitation of sensitive personal information, impersonation of trusted institutions, urgency cues, and the use of suspicious or unverifiable communication channels. These indicators were used to ensure consistency in scenario design and evaluation across all fraud categories.

To ensure consistency and theoretical grounding in the design of fraud scenarios, each stimulus message incorporated one or more commonly observed deceptive cues identified in prior digital fraud research. These cues represent typical manipulation strategies used by scammers to influence individuals' decision-making processes. The cues included urgency signals, authority impersonation, reward promises, emotional manipulation, and suspicious hyperlinks.

Each fraud scenario was systematically constructed by embedding these deceptive cues within realistic communication formats such as emails, text messages, and social media notifications. This approach allowed the experimental materials to simulate authentic scam characteristics while maintaining control over the key manipulative elements.

From a Person–Task Fit perspective, these deceptive cues represent critical task characteristics that influence the cognitive demands of fraud detection. By providing explanations for these cues, the AI-assisted system evaluated in this study aims to help users better interpret suspicious signals and improve their alignment with the task requirements of identifying fraudulent communications.

The six fraud categories were selected through a review of common fraud typologies reported in prior research and public-facing fraud-prevention sources, with attention to both international and local relevance. Internationally, consumer-protection data and scam reports commonly identify categories such as imposter scams, online shopping fraud, investment fraud, phishing or account-security scams, business or employment-related scams, and social or relationship-based scams. For example, recent Federal Trade Commission data identify imposter scams as the most commonly reported fraud category in the United States, followed by online shopping issues, business and job opportunities, investment-related reports, and internet services (Federal Trade Commission, 2025). In the Hong Kong context, police and anti-deception statistics similarly highlight e-shopping fraud, online investment fraud, online employment fraud, phishing scams, and telephone deception involving impersonation tactics as major categories of deception (Anti-Deception Coordination Centre, 2026; Hong Kong Police Force, 2026). In mainland China, public security and financial-regulatory sources also identify high-frequency or emerging forms of telecom and online fraud, including false shopping or service fraud, impersonation of customer service or official institutions, false investment and financial-management fraud, phishing-like credential theft, and AI-enabled impersonation or relationship-based deception (Ministry of Public Security of the People's Republic of China, 2024; National Financial Regulatory Administration, 2024). Based on these overlapping sources, we grouped the scenarios into six broad categories that were suitable for the present experimental design and developed to reflect common categories of digital fraud that are frequently encountered in everyday online communication: *financial/payment fraud*, *cyber/technical fraud*, *social relationship scams*, *consumer fraud*, *government or institutional impersonation fraud*, and *social media scams*. Each category was operationalized through realistic stimuli presented in familiar communication formats, including emails, text messages, app notifications, and chat-style messages. Across all categories, the scenarios were designed to preserve ecological validity while ensuring participant safety; thus, no scenario required actual money transfer, disclosure of real personal

Table 4
Fraud scenario categories, operational definitions, deceptive cues, example stimuli, and classification criteria.

Fraud scenario type	Operational definition	Key deceptive cues	Example scenarios (Stimuli Examples)	Fraud classification criteria
Financial/Payment Fraud	Messages attempting to obtain money or financial credentials by impersonating financial institutions or presenting urgent financial issues.	• Urgency cues (e.g., "act immediately") • Financial incentives or threats • Authority impersonation (banks, payment platforms) • Requests for financial credentials	1. Fake bank alert claiming the account is frozen and requiring immediate verification. 2. Investment opportunity promising unusually high returns. 3. Payment request to avoid account suspension. 4. Message asking for card details to process a refund.	The message is classified as fraudulent if it: • Requests financial transfers or payment credentials. • Uses urgency related to financial consequences. • Comes from unverifiable financial institutions. • Includes suspicious payment channels.
Cyber/Technical Fraud	Digital messages designed to obtain login credentials or personal information through phishing links or fake security alerts.	• Suspicious hyperlinks • Security warnings or threats • Requests for login credentials • Fake system notifications	1. Email claiming unusual login activity requiring password reset. 2. Security alert requesting identity verification via link. 3. Technical support message stating the device has a virus. 4. Fake platform notification requesting login confirmation.	Classified as fraud if it: • Contains suspicious links or fake login pages. • Requests passwords or verification codes. • Uses urgent security warnings. • Sender identity cannot be verified.
Social Relationship Scams	Messages exploiting personal relationships or emotional trust to request assistance, money, or sensitive information.	• Emotional manipulation (sympathy or urgency) • Trust exploitation • Emergency narratives • Requests for quick assistance	1. A "friend" asking for urgent financial help. 2. A new online contact claiming an emergency situation. 3. Message pretending to be a relative needing assistance. 4. A social contact requesting a verification code.	Classified as fraud if it: • Requests money or sensitive information from unverified contacts. • Uses emotional pressure or emergencies. • Requires immediate action without verification. • Uses newly created or suspicious accounts.
Consumer Fraud	Fraudulent commercial messages mimicking legitimate online shopping, delivery, or refund notifications.	• Fake promotional offers • Suspicious payment links • Order verification requests • Delivery or refund urgency	1. Fake delivery message requesting payment for shipping. 2. Refund notification requiring bank details. 3. Promotional offer requiring account verification. 4. Message asking to confirm order information via external link.	Classified as fraud if it: • Directs users to unofficial payment pages. • Requests financial details for refunds or delivery fees. • Uses suspicious domains or links. • Offers unrealistic discounts or rewards.
Government/Institutional Impersonation Fraud	Messages impersonating government agencies or official institutions to pressure users into providing information or payments.	• Authority cues (government, police, tax office) • Legal threats or penalties • Urgency language • Requests for identity verification	1. Fake tax notice requesting immediate payment. 2. Message claiming unpaid fines requiring identity verification. 3. Email pretending to be from a government agency requesting personal data. 4. Warning about legal consequences if payment is not made immediately.	Classified as fraud if it: • Claims authority without verifiable official channels. • Requests payment or personal information through unofficial links. • Uses legal threats to create urgency. • Sender identity does not match official institutions.
Social Media Scams	Fraudulent messages delivered through social media platforms attempting to redirect users to external sites or solicit sensitive information.	• Prize or reward promises • Suspicious private messages • Links to external websites • Newly created or fake accounts	1. Message claiming the user has won a prize. 2. Investment promotion from an unknown account. 3. Private message asking the user to verify account information. 4. Social media promotion requiring login through an external link.	Classified as fraud if it: • Originates from suspicious or newly created accounts. • Redirects users to external sites requesting credentials. • Promises unrealistic rewards. • Requests sensitive information via private messaging.

information, or engagement with genuine external websites. A full overview of the fraud categories, their operational definitions, key deceptive cues, example stimuli, and classification criteria is provided in Table 4.

Although the six fraud categories were held constant across age groups, some scenario instantiations were age-adapted to improve ecological validity. The purpose of this adaptation was to ensure that the communication format, social framing, and credibility cues were realistic for each age group while preserving functional equivalence within each fraud category. For example, within social media fraud, scenarios for the younger group more often involved online relationship building, shared interests, gaming or hobby communities, and small transfers framed as tokens of trust or platform-based support. Scenarios for the older group more often involved a sender claiming to be a former friend, old classmate, retired colleague, or other familiar contact reappearing through digital platforms and then requesting assistance or a small transfer. Similarly, within consumer fraud, scenarios for the younger group more often involved discounts, prepaid access,

Table 5
Deceptive Cue Categories Used in Fraud Scenarios.

Deceptive Cue	Description	Example
Urgency Cue	Messages create time pressure to force quick decisions and reduce users' ability to critically evaluate the information.	"Your account will be suspended within 24 h unless you verify your information immediately."
Authority Cue	Messages impersonate trusted institutions such as banks, government agencies, or well-known companies to increase perceived legitimacy.	"This is a notification from the National Tax Authority regarding unpaid taxes."
Reward Cue	Messages promise financial rewards, prizes, or exclusive opportunities in order to attract attention and motivate user interaction.	"Congratulations! You have been selected to receive a $500 reward."
Emotional Manipulation	Messages exploit emotional responses such as sympathy, fear, or trust to encourage compliance.	"I am currently in an emergency situation and urgently need your help."
Suspicious Link	Messages include hyperlinks directing users to external websites designed to collect credentials or sensitive information.	"Please verify your account by clicking the link below."

Table 6
AI risk explanation and decision support output across fraud types.

Fraud type	AI risk level	AI explanation	Suggested action
Financial/Payment Fraud	High	The message contains urgent payment language, requests financial credentials, and cannot be verified through an official financial channel. These cues are commonly associated with payment scams and fraudulent refund or transfer requests.	Do not transfer money or provide card details. Verify the request directly through the bank's or payment platform's official website or customer service channel.
Cyber/Technical Fraud	High	The message includes a suspicious login or verification link, claims account compromise, and pressures the user to act immediately. Such combinations are typical indicators of phishing and technical-support fraud.	Do not click the link or provide login credentials. Access the account only through the official app or website and independently confirm whether any security issue exists.
Social Relationship Scams	Moderate to High	The message relies on emotional urgency, requests immediate help, and comes from an identity that has not been independently verified. This pattern is consistent with scams exploiting trust, sympathy, or personal relationships.	Do not send money or share sensitive information until the sender's identity is confirmed through a trusted and independent communication channel.
Consumer Fraud	Moderate to High	The message presents an order, refund, or delivery problem and asks the user to confirm payment or account information through an external link. The domain, payment request, or promotional content appears inconsistent with legitimate commercial communication.	Do not click the embedded link or submit payment information. Check the order or refund status directly on the retailer's official website or app.
Government/Institutional Impersonation Fraud	High	The message claims to originate from an official institution, uses threat-based language about penalties or legal consequences, and requests immediate payment or identity verification through unofficial contact routes. These are strong indicators of impersonation fraud.	Do not respond or make payment. Contact the institution through its official public contact information to verify whether the notice is authentic.
Social Media Scams	Moderate to High	The message comes from a suspicious or unfamiliar account, promises rewards or opportunities, and redirects the user to an external site or asks for account verification. These features are commonly observed in fraudulent social media messages.	Do not engage with the link or message content. Report or block the account if appropriate, and verify any promotion or account issue only through the platform's official interface.

group-buying, or small deposits, whereas scenarios for the older group more often involved routine service fees, delivery issues, or small charges presented as necessary for verification or processing. Within government or institutional impersonation fraud, scenarios for the older group more often emphasized official procedure, politeness, in-person inspection, or minor administrative fees, whereas scenarios for the younger group more often involved digitally mediated variants such as online verification notices or account-related institutional alerts. These adaptations were designed to preserve ecological validity while maintaining category-level comparability across age groups.

To enhance internal consistency in the AI-assisted condition, the decision support outputs were standardized across fraud categories. For each type of scenario, the AI system produced a structured combination of risk level, explanatory rationale, and recommended user response. Table 6 presents representative examples of these AI-generated outputs across the six fraud categories. This standardization ensured that any observed intervention effect could be attributed to the presence of AI-supported interpretation and guidance rather than to uncontrolled differences in feedback format.

All experimental materials were pilot-tested with a separate group of participants prior to the main study. The pilot procedure examined the perceived realism, credibility, clarity, and difficulty of the scenarios, as well as the comprehensibility of the AI-generated explanations. Based on participant feedback, minor revisions were made to wording, message formatting, and cue salience to improve authenticity while maintaining comparability across fraud types.

3.4. Procedure

The study was implemented over a 10-day period and consisted of three sequential phases: *baseline assessment*, *fraud-exposure tasks*, and *post-study evaluation*. Because the baseline questionnaire has already

Table 7
Overview of the study procedure.

Phase	Timing	Main activities	Data collected
Baseline assessment	Day 0	Participant orientation, informed consent, baseline questionnaire, and initial familiarization with the study platform. Participants in the AI-assisted condition also received a brief introduction to the AI support interface.	Demographic characteristics, digital literacy, prior fraud exposure, previous scam-related harm, risk perception, confidence in detecting suspicious messages, and digital usage habits.
Fraud-exposure phase	Days 1–10	Participants received simulated fraud messages at irregular intervals through the study platform. Each participant encountered multiple fraud scenarios across the six fraud categories. Control-group participants made judgments independently, whereas AI-group participants could consult AI-generated risk assessment, cue explanations, and recommended actions before finalizing their responses.	Fraud judgment, intended behavioral response, confidence rating, response time, and, in the AI-assisted condition, system consultation behavior.
Post-study evaluation	Day 10	Participants completed a follow-up questionnaire after the exposure phase. A subset of participants was then invited to take part in semi-structured interviews regarding their decision-making experiences and perceptions of the AI tool.	Perceived usefulness of the AI system, trust in AI-generated advice, perceived realism of the scenarios, self-reported confidence in fraud detection, and qualitative reflections on decision-making processes.

been described in the participant assessment section, only a brief summary is provided here, with emphasis placed on how the assessment was integrated into the overall experimental workflow. Table 7 presents an overview of the study procedure.

Prior to the beginning of the experimental exposure phase, all participants completed the baseline assessment described above. This stage took place on **Day 0** and served two purposes: to document participants' initial levels of fraud-relevant knowledge and experience, and to confirm that the experimental groups were broadly comparable before the intervention. As noted earlier, the baseline questionnaire assessed demographic characteristics, digital literacy, prior fraud exposure, previous scam-related harm, risk perception, confidence in detecting suspicious messages, and everyday digital usage habits, in order to characterize participants' pre-intervention profiles and evaluate baseline comparability across groups. In addition, participants received a brief orientation to the study platform and the general task format. For those assigned to the AI-assisted condition, this orientation also included a short demonstration of how to access and interpret the AI support output. No fraud-specific feedback or training was provided at this stage, so as to avoid influencing subsequent responses.

The fraud-exposure phase was conducted from **Day 1 to Day 10**. During this period, participants accessed a web-based study platform using their own smartphones or computers. The platform was developed for this study so that simulated fraud scenarios could be presented in a controlled format and participants' response time could be automatically recorded from the moment a scenario was opened to the moment a final response was submitted. Participants used their own devices in order to preserve their everyday digital habits and approximate real-world usage conditions, since message evaluation may differ across smartphones, computers, and individual interaction routines. At irregular intervals, participants were prompted to access the platform and complete newly available fraud-detection tasks, thereby approximating the unpredictability of real-world digital fraud exposure while maintaining experimental control. Across the exposure period, each participant encountered multiple fraud scenarios representing the six fraud categories described in the previous section: financial/payment, cyber/technical, social-relationship, consumer, government/institutional impersonation, and social-media fraud. The scenarios were displayed within the platform using realistic simulated interface formats, such as email-style messages, text-message screens, app-notification-style prompts, and chat-style messages.

For each scenario, participants were asked to make a judgment regarding whether the communication appeared fraudulent, indicate whether they would be willing to interact with it (e.g., click a link, reply, provide information, or otherwise comply), and rate their confidence in that decision. Participants in the *control condition* completed these tasks without any form of technological assistance and relied solely on their own judgment. Participants in the *AI-assisted condition*, by contrast, were able to consult the AI decision support system before submitting their final response. As described in the previous subsection, the AI output included a risk assessment, a brief explanation of suspicious cues, and an action recommendation. This structure allowed us to examine whether access to AI-supported interpretation changed participants' fraud judgments and intended behavioral responses under ecologically valid task conditions.

At the end of the exposure period on **Day 10**, participants completed a post-study evaluation. This follow-up questionnaire assessed perceived usefulness of the AI system, trust in AI-generated advice, perceived realism of the fraud scenarios, and self-reported confidence in fraud detection. These measures were included to capture participants' subjective experiences of the intervention and to complement the behavioral outcomes collected during the exposure phase. In addition, a subset of participants from both age groups and both experimental conditions was invited to participate in semi-structured interviews. The interview subset was selected to include participants from both age groups and both experimental conditions, so that the qualitative data could capture variation in experiences with and without AI support. These interviews explored how participants interpreted suspicious messages, what cues they relied on when making decisions, how they experienced the AI support tool when available, and whether they felt that the intervention influenced their confidence or decision strategies. The qualitative component was included to provide a richer account of participants' decision-making processes and to support interpretation of the quantitative findings.

3.5. Measures

The study collected a set of behavioral, cognitive, and intervention-related measures to evaluate participants' responses to the simulated fraud scenarios. Unless otherwise specified, these measures were recorded at the scenario level during the exposure phase and were later aggregated at the participant level for analysis where appropriate. Together, they were intended to capture participants' ability to recognize fraudulent communications, their willingness to engage in risky actions, the confidence and effort involved in their decisions, and, for those in the AI-assisted condition, the extent to which the AI system influenced their responses.

- **Fraud detection accuracy.** Fraud detection accuracy referred to participants' ability to correctly identify a presented communication as fraudulent. For each scenario, responses were coded as accurate (1) if the participant correctly classified the message as a scam and inaccurate (0) otherwise. Because all experimental stimuli were constructed as simulated fraud scenarios, this measure reflected participants' success in recognizing deceptive communications under ecologically valid conditions. In addition to overall performance, fraud detection accuracy was also examined separately across the six fraud categories in order to assess whether some scam types were more difficult to recognize than others.
- **Risky behavioral intention.** Risky behavioral intention captured participants' stated willingness to engage with a suspicious communication in ways that could increase their vulnerability to fraud. Depending on the scenario, this included actions such as clicking a link, replying to a message, providing information, continuing an interaction, making a payment, or otherwise complying with the request. In the final study, this outcome was measured using a five-point Likert-type item tailored to the scenario-specific risky action, ranging from 1 = definitely would not do this to 5 = definitely would do this. Higher scores therefore indicated greater willingness to engage in potentially unsafe actions. A Likert-type format was used rather than a binary yes/no response because the scenarios differed in the type and perceived severity of the requested action, and the study aimed to capture graded levels of behavioral intention rather than only final compliance or non-compliance.
- **Decision confidence.** Decision confidence captured participants' self-reported certainty in their fraud-related judgment for each scenario. After making a decision, participants rated how confident they were that their evaluation was correct. This measure was included to assess participants' subjective certainty and to distinguish between decisions made confidently and those made with hesitation or uncertainty. It was also useful for interpreting whether AI assistance improved not only decision quality but also the degree to which participants felt assured about their judgments.
- **Decision time.** Decision time was defined as the time required for participants to evaluate each scenario and submit their response. This measure was automatically recorded by the study platform from the moment the scenario was opened to the moment the decision was finalized. Decision time was included as an indicator of cognitive effort and processing demand. It also provided insight into whether AI assistance reduced decision burden by facilitating faster judgments or instead encouraged more deliberative processing of suspicious messages.
- **Treatment check and intervention engagement.** For participants in the AI-assisted condition, treatment exposure was assessed using system-recorded interaction logs. Specifically, *AI consultation frequency* referred to the proportion of scenarios in which participants opened and viewed the AI support output during the exposure phase. The platform also recorded whether participants viewed only the risk assessment or additionally accessed the explanatory rationale and recommended action. These measures were used to verify the extent to which participants assigned to the AI-assisted condition were actually exposed to and engaged with the intervention. Participants in the control condition did not have access to the AI decision-support functions.
- **AI-assisted decision revision.** Separately from the treatment check, *AI-assisted decision improvement* captured whether participants revised their initial judgment after consulting the AI system and whether that revision resulted in a safer or more accurate final decision. Improvement was coded positively when the final decision became more accurate or less risky after AI consultation, negatively when the final decision became less accurate or more risky, and neutrally when no substantive change occurred. This measure was treated as an intervention-related process outcome rather than as an indicator of treatment exposure.
- **Post-study evaluative measures.** At the end of the exposure phase, participants completed a follow-up questionnaire assessing several evaluative constructs relevant to the intervention. These included *perceived usefulness of the AI system*, *trust in AI-generated advice*, *perceived realism of the fraud scenarios*, and *self-reported confidence in fraud detection*. Although these measures were collected after the main exposure phase, they formed part of the overall measurement framework because they provided important contextual information for interpreting participants' behavioral responses and experiences with the AI-assisted decision support system.

All scenario-level measures were collected through the study platform immediately after exposure to each simulated fraud message. Fraud detection accuracy was obtained from participants' binary classification of whether the presented communication was fraudulent, with correct identification coded as 1 and incorrect identification coded as 0. Risky behavioral intention was derived from participants' reported willingness to engage in scenario-relevant actions, such as clicking a link, replying to the sender, providing information, making a payment, or continuing the interaction. These responses were recorded using five-point Likert-type ratings, with higher scores indicating greater willingness to engage in potentially unsafe actions. Decision confidence was measured by asking participants to rate how certain they were that their judgment was correct immediately after each scenario, using a Likert-type confidence scale. Decision time was automatically recorded by the platform as the interval between opening the scenario and submitting the final response.

For participants in the AI-assisted condition, the platform additionally logged AI interaction behavior. Specifically, the system recorded whether participants opened the AI support output, whether they viewed only the risk label or also examined the explanatory rationale and recommended action, and whether their final decision changed after consultation. AI-assisted decision improvement was derived by comparing participants' initial judgment with their final judgment after viewing the AI output; decision changes were coded according to whether they led to a safer or more accurate outcome, a less safe or less accurate outcome, or no substantive change. At the end of the 10-day exposure period, participants completed a follow-up questionnaire assessing perceived usefulness of the AI system, trust in AI-generated advice, perceived realism of the fraud scenarios, and self-reported confidence in fraud detection, with responses recorded on Likert-type scales and aggregated at the construct level where appropriate.

3.6. Data analysis

The data analysis strategy was designed to match the repeated-exposure and mixed-methods nature of the study. Because each participant responded to multiple fraud scenarios across the 10-day study period, the dataset had a nested structure in which scenario-level observations were clustered within participants. To account for this non-independence, the primary quantitative analyses were conducted using multilevel regression models. This approach allowed us to examine both between-group differences (AI-assisted vs. control; younger vs. older groups) and within-person variation across fraud types while appropriately modeling repeated observations from the same participant. Before testing the substantive research questions, treatment adherence in the AI-assisted condition was examined using system-recorded interaction logs. Specifically, AI consultation frequency and engagement with the explanatory and recommended-action components were summarized to assess the extent to which participants actually accessed and used the assigned intervention. These treatment-check measures were analyzed separately from the study outcomes. Fraud detection

accuracy and risky behavioral intention were treated as the primary behavioral outcomes, decision confidence and decision time as secondary outcomes, and AI-assisted decision revision as an intervention-related process outcome.

The primary analyses focused on the two main behavioral outcomes of the study: *fraud detection accuracy* and *risky behavioral intention*. Fraud detection accuracy was coded as a binary outcome indicating whether a participant correctly identified a presented communication as fraudulent. Accordingly, this variable was analyzed using multilevel logistic regression. Risky behavioral intention, which reflected participants' willingness to engage in scenario-relevant risky actions, was measured using five-point Likert-type responses in the final dataset and was therefore analyzed using linear mixed-effects models. This format was selected because risky responses varied across scenarios, including clicking links, replying to messages, providing information, making payments, or continuing interaction. A graded Likert-type measure allowed us to capture variation in willingness to engage in these potentially unsafe actions, rather than reducing responses to a binary comply/do-not-comply distinction. In both cases, the fixed effects included *AI condition* (AI-assisted vs. control), *age group* (younger vs. older groups), and *fraud type* (the six scam categories), together with their interaction terms. Participant was entered as a random intercept to account for stable between-person differences in baseline fraud susceptibility and decision style. Baseline covariates, including digital literacy, prior fraud exposure, perceived fraud risk, and daily digital use, were included in adjusted models where appropriate. The general model structure for the primary analyses was as follows:

$$Outcome \sim AICondition \times AgeGroup \times FraudType + BaselineCovariates + (1 \mid Participant)$$

where *Outcome* referred to either fraud detection accuracy or risky behavioral intention. This modeling strategy enabled us to test the overall effectiveness of AI-assisted support, age-related differences in fraud-related decision making, and whether the intervention effect varied across scam categories.

Secondary analyses were conducted for *decision confidence* and *decision time*. Because these variables were treated as continuous outcomes, they were analyzed using linear mixed-effects models with the same fixed-effect structure as the primary models. Decision confidence was analyzed to examine whether AI-assisted support increased participants' subjective certainty in their judgments, whereas decision time was analyzed as an indicator of cognitive effort and deliberation. These analyses were intended to determine whether AI support affected not only decision quality, but also the subjective and temporal characteristics of fraud-related decision making.

To further examine differences across the six scam categories, fraud-type-specific descriptive and comparative analyses were also conducted. Specifically, fraud detection accuracy and risky behavioral intention were summarized separately for financial/payment fraud, cyber/technical fraud, social relationship scams, consumer fraud, government or institutional impersonation fraud, and social media scams. These analyses were used to identify which scam types were most difficult to detect, which elicited the highest level of risky compliance, and whether the effectiveness of AI-assisted support was stronger for some fraud categories than for others.

For participants in the *AI-assisted condition*, treatment–adherence and intervention-related process measures were analyzed separately. AI consultation frequency and engagement with the explanatory and recommended-action components were summarized as treatment-adherence indicators, as described above. Separately, *AI-assisted decision revision* was examined by comparing participants' initial judgments with their final judgments after consultation with the AI output. This process analysis was used to characterize whether decision revisions following AI consultation were associated with safer or more accurate final responses.

The *post-study evaluative measures*, including perceived usefulness of the AI system, trust in AI-generated advice, perceived realism of the fraud scenarios, and self-reported confidence in fraud detection, were analyzed using descriptive statistics and between-group comparisons. These analyses were intended to complement the scenario-level behavioral data by capturing participants' overall perceptions of the intervention and the ecological plausibility of the study materials.

Statistical assumption checks. Prior to inferential analysis, the assumptions associated with each statistical procedure were examined. For the linear mixed-effects models, the distribution of model residuals was evaluated using Q–Q plots and Shapiro–Wilk tests, while residual-versus-fitted plots were inspected for evidence of heteroscedasticity. Standardized residuals were also examined to identify extreme observations. Model convergence and multicollinearity among the fixed effects were assessed before interpreting the model estimates. The diagnostic results indicated no substantial violations of residual normality or homoscedasticity, and all fitted models converged successfully without evidence of problematic multicollinearity. For the multilevel logistic regression models, normality of the binary outcome was not assumed; instead, model convergence, dispersion, and residual diagnostics were examined to assess model adequacy. No substantial evidence of model misspecification was observed. For one-way ANOVA and independent-samples and one-sample *t*-tests, distributional assumptions were assessed using Q–Q plots and Shapiro–Wilk tests. Homogeneity of variance for between-group comparisons was evaluated using Levene's test. No substantial violations were identified for the reported parametric comparisons. Finally, expected cell frequencies were examined for categorical comparisons. Fisher's exact test, or its extension for larger contingency tables, was used instead of Pearson's χ^2 test when expected frequencies were insufficient for the asymptotic approximation.

Finally, the qualitative interview data were analyzed using thematic analysis. The analysis focused primarily on semantic coding, meaning that codes were developed from the explicit content of participants' accounts rather than from inferred latent meanings. All interview transcripts were first read in full by two members of the research team to gain familiarity with the data. The two coders then independently coded an initial subset of transcripts and developed preliminary codes related to participants' fraud-detection strategies, the cues they relied on when making judgments, their experiences of uncertainty, and their perceptions of the usefulness and trustworthiness of the AI system. Inter-coder agreement for this initial coding stage was high, with Cohen's $\kappa = 0.89$, indicating strong consistency before discussion and reconciliation. After this initial coding stage, the coders held a theming meeting to compare codes, resolve discrepancies through discussion, and develop a shared coding framework. The remaining transcripts were then coded using this framework, with new codes added where necessary. A second theming meeting was held to review the final code structure, merge overlapping codes, and organize the codes into broader themes. Disagreements were resolved through discussion between the two coders, and a third member of the research team reviewed the final themes to check whether they were clearly supported by the interview data. The qualitative analysis required approximately 36 person-hours in total, including transcript familiarization, independent coding, theming meetings, discrepancy resolution, and final theme review. The qualitative findings were used to enrich the interpretation of the quantitative results, particularly in understanding why AI support was helpful or unhelpful in different fraud contexts and how participants experienced the intervention under ecologically valid conditions.

4. Results

4.1. Treatment check

Treatment adherence was examined before testing the substantive research questions. By design, participants in the control condition

Table 8
Descriptive results for RQ1 by age group and decision-support condition.

Measure	YA-Control	YA-AI	OA-Control	OA-AI
Fraud detection accuracy, Mean (SD)	0.71 (0.10)	0.86 (0.08)	0.62 (0.12)	0.79 (0.10)
Risky behavioral intention, Mean (SD)	2.48 (0.42)	1.89 (0.39)	3.01 (0.47)	2.34 (0.44)
Decision confidence, Mean (SD)	3.42 (0.51)	3.78 (0.47)	3.26 (0.49)	3.61 (0.46)
Decision time (seconds), Mean (SD)	12.9 (4.4)	21.6 (6.0)	19.8 (6.7)	30.5 (7.5)

Note. YA = younger group; OA = older group. Higher values for risky behavioral intention indicate greater willingness to engage in potentially unsafe actions. Risky behavioral intention was measured on a five-point Likert-type scale, with higher values indicating greater willingness to engage in potentially unsafe actions.

completed the fraud-detection tasks without access to the AI decision-support functions, whereas participants assigned to the AI-assisted condition could access the system-generated risk assessment, explanation, and recommended action. System interaction logs showed that participants in both age groups engaged with the assigned intervention. Younger participants consulted the AI support for an average of 84% of available scenarios ($SD = 0.11$), whereas older participants consulted it for 63% ($SD = 0.14$). Among scenarios in which the AI system was consulted, the recommended-action component was viewed in 56% ($SD = 0.18$) of consultations by younger participants and 78% ($SD = 0.15$) by older participants. These usage patterns indicate substantial exposure to the AI intervention in both age groups, while also showing age-related differences in how participants engaged with its components. These measures were used solely as indicators of treatment adherence and were analyzed separately from the behavioral and intervention-related outcomes reported below.

4.2. Effects of AI-assisted decision support on fraud detection and risky behavioral intention (RQ1)

In this section, we examined whether access to the AI-assisted decision support system improved participants' fraud detection accuracy and reduced risky behavioral intention relative to unaided judgment. Because participants completed repeated fraud-detection tasks across the 10-day online lab-in-the-field experiment, the analyses combined descriptive comparisons at the group level with multilevel regression models at the scenario level. In addition to the two primary outcomes, we also examined decision confidence, decision time, and intervention-related decision-revision outcomes to better understand the behavioral and process-level effects of the intervention across age groups.

Table 8 presents the descriptive results for the four experimental groups. Overall, participants in the AI-assisted condition showed higher fraud detection accuracy and lower risky behavioral intention than those in the control condition in both age groups. Younger group in the AI-assisted condition achieved the highest accuracy, whereas older group in the control condition showed the lowest accuracy and the highest level of risky behavioral intention. Older group also required more time to make decisions than younger group across both conditions. Participants in the AI-assisted condition took longer to complete each scenario than those in the control condition.

To formally test these patterns, multilevel regression models were estimated with participant as a random intercept and fraud type included as a repeated within-person factor. The results are summarized in Table 9. For *fraud detection accuracy*, AI-assisted decision support showed a significant positive effect, indicating that participants with access to AI were substantially more likely to correctly identify fraudulent communications than those in the control condition. Participants in the older group showed significantly lower detection accuracy than the younger group overall. For *risky behavioral intention*, AI-assisted support was associated with a significant reduction in participants' willingness to comply with suspicious requests, whereas the older group showed significantly higher risky behavioral intention than the younger group, suggesting that the intervention was beneficial in both age groups, even though the older group remained relatively more vulnerable overall.

The descriptive and inferential results consistently indicate that AI-assisted support improved performance on the core fraud-detection task. Compared with unaided participants, those in the AI-assisted condition were more accurate in identifying scams and less willing to engage in risky actions. At the same time, the older group remained more vulnerable overall, as reflected in both lower detection accuracy and higher risky behavioral intention, and took significantly longer to make decisions than the younger group. Additional analyses within the AI-assisted condition examined decision revisions following AI consultation. As shown in Table 10, the proportion of scenarios involving a safer decision revision was significantly higher among older participants than among younger participants. Among decisions that were revised after AI consultation, successful avoidance rates were high in both age groups and did not differ significantly between them. These process outcomes were analyzed separately from the treatment-adherence measures reported in the Treatment Check section.

Taken together, the results for RQ1 provide converging evidence that AI-assisted decision support improved fraud-related decision making in the simulated digital fraud scenarios. The intervention significantly increased fraud detection accuracy and reduced risky behavioral intention relative to unaided human judgment. Older participants nevertheless showed greater overall vulnerability and longer decision times. Within the AI-assisted condition, decision revisions following AI consultation were frequently associated with safer final responses in both age groups, providing additional process-level evidence on how participants used the intervention during fraud-related decision making.

Findings for RQ1: AI-assisted decision support significantly improved fraud detection accuracy and reduced risky behavioral intention relative to unaided judgment. Although older group showed greater vulnerability and longer decision times, AI-supported decision revisions in both age groups were largely associated with safer final responses.

4.3. Age-related vulnerability patterns across scam types and the moderating role of AI-assisted support (RQ2)

In this section, we examined age-related fraud vulnerability by comparing younger and older group across the six fraud categories and by assessing whether AI-assisted support reduced the proportion of risky or deceptive responses within each age group. Table 11 summarizes the number of deceptive or risky responses out of the total number of scenario exposures across the four experimental groups. Across the 10-day study period, each group encountered an equivalent number of exposures within each fraud category, allowing direct comparison of scam-specific vulnerability rates. In the control condition, the younger

Table 9
Multilevel regression results for the primary and secondary RQ1 outcomes with effect sizes.

Predictor	Estimate	SE	Test statistic	p	Effect size
Fraud detection accuracy (multilevel logistic regression)					
AI condition (AI vs. control)	1.08	0.24	$z = 4.50$	< 0.001	OR = 2.94
Age group (older vs. younger)	−0.76	0.26	$z = -2.92$	0.003	OR = 0.47
AI × Age group	0.19	0.21	$z = 0.90$	0.365	OR = 1.21
Fraud type	–	–	$\chi^2(5) = 18.42$	0.002	$w = 0.15$
Risky behavioral intention (linear mixed-effects model)					
AI condition (AI vs. control)	−0.56	0.09	$t = -6.22$	< 0.001	$r = -.22$
Age group (older vs. younger)	0.43	0.10	$t = 4.30$	< 0.001	$r = 0.15$
AI × Age group	−0.08	0.07	$t = -1.07$	0.286	$r = -.04$
Fraud type	–	–	$F(5, 792) = 7.91$	< 0.001	$\eta_p^2 = 0.048$
Decision confidence (linear mixed-effects model)					
AI condition (AI vs. control)	0.31	0.11	$t = 2.82$	0.005	$r = 0.10$
Age group (older vs. younger)	−0.14	0.11	$t = -1.27$	0.205	$r = -.05$
AI × Age group	0.03	0.08	$t = 0.38$	0.706	$r = 0.01$
Decision time (linear mixed-effects model)					
AI condition (AI vs. control)	9.70	1.02	$t = 9.51$	< 0.001	$r = 0.32$
Age group (older vs. younger)	7.90	1.11	$t = 7.12$	< 0.001	$r = 0.25$
AI × Age group	2.00	1.44	$t = 1.39$	0.165	$r = 0.05$

Note. Positive coefficients for fraud detection accuracy indicate better performance; negative coefficients for risky behavioral intention indicate lower vulnerability. For logistic regression, odds ratios (ORs) are reported as effect sizes; for linear mixed-effects models, approximate partial r values are reported; for omnibus fraud-type effects, Cohen's w or partial eta-squared (η_p^2) is reported. **Practical significance:** AI support nearly tripled the odds of correct fraud detection, reduced risky behavioral intention by 0.56 points on a five-point scale, modestly increased decision confidence, and increased decision time by approximately 9.70 s, suggesting safer and more deliberative fraud-related decision making.

Table 10
AI-assisted decision revision outcomes within the AI-assisted condition.

Measure	Younger group	Older group	Statistic	p
AI-assisted decision improvement (proportion of scenarios with safer revision), Mean (SD)	0.22 (0.09)	0.31 (0.11)	$t(34) = -2.43$	0.020
Successful avoidance after AI-based revision, n (%)	31/35 (88.6%)	44/48 (91.7%)	Fisher's exact test	0.716

Note. AI-assisted decision improvement refers to the proportion of scenarios in which consultation with the AI system was followed by a safer or more accurate final decision. Successful avoidance after AI-based revision refers to the proportion of revised decisions that resulted in a safer final response, such as refusing to click a link, declining to reply, or withholding personal or financial information. Fisher's exact test was used for the successful-avoidance comparison because of small expected cell frequencies.

group showed the highest deceptive-response rates for social-media fraud and consumer fraud, whereas the older group showed the highest deceptive-response rates for government/institutional impersonation fraud and financial/payment fraud. In the AI-assisted condition, deceptive-response rates were lower in both age groups across all fraud categories.

A closer inspection of the scenarios that were most frequently *misclassified as safe* or elicited *willingness to comply with the fraudulent request* revealed a common pattern across age groups. Many of these scenarios involved relatively **small monetary amounts** and low-threshold requests, such as paying a modest deposit, scanning a QR code to join a promotional group after a small payment, or responding to a supposed government or inspection-related visit that requested a minor service fee. Compared with scams involving large and obviously risky transfers, these low-cost requests appeared to reduce participants' sense of danger and increase their willingness to "try" or comply, especially when the message also included cues of convenience, legitimacy, or limited-time benefit. For younger group, this pattern was particularly visible in social-media- and consumption-related scenarios, where small payments were framed as part of promotions, discounts, or platform-based opportunities. For older group, it was more visible in authority-based scenarios, such as impersonation messages or in-person inspection requests, where small fees were presented as routine, necessary, or officially required. Importantly, the AI-assisted condition reduced deceptive responses in both age groups, with particularly visible reductions in the scam categories that produced the highest baseline vulnerability for each group. These findings suggest that age-related fraud vulnerability is shaped not only by scam type, but also by how fraudulent requests are framed as *low-cost*, *reasonable*, or *routine*, thereby lowering suspicion and increasing compliance. In this sense, the findings indicate that the persuasive power of these scams may lie less in the absolute amount requested than in the perception that the request is minor, manageable, and therefore not worth scrutinizing in depth.

To formally test these patterns, a multilevel logistic regression model was estimated with deceptive response as the binary outcome. The model included AI condition, age group, fraud type, and their interaction terms, with participant entered as a random intercept. The results are summarized in Table 12. The omnibus interaction between age group and fraud type was statistically significant, indicating that younger and older groups showed different fraud-category-specific vulnerability profiles. The interaction between AI condition and fraud type approached, but did not reach, conventional statistical significance, suggesting a possible tendency for the effect of AI-assisted support to vary across fraud categories. Similarly, the omnibus interaction between age group and fraud type was statistically significant, indicating that deceptive-response likelihood varied across fraud categories differently for the two age groups. The interaction between AI condition and fraud type approached, but did not reach, conventional statistical significance. The three-way interaction between AI condition, age group, and fraud type was also suggestive but not statistically significant at the conventional .05 level. Therefore, the fraud-category-specific comparisons reported below should be interpreted as descriptive patterns rather than confirmatory evidence of differential intervention effects by age group and fraud category.

To clarify the interaction patterns, follow-up comparisons were conducted within each age group. Among younger group, the highest deceptive-response rates in the control condition were observed for *social media scams* (38.9%) and *consumer fraud* (36.1%). In the AI-assisted

Table 11
Deceptive or risky responses out of total exposures by age group, AI condition, and fraud type.

Fraud type	YA-Control	YA-AI	OA-Control	OA-AI
Financial/Payment Fraud	7/36 (19.4%)	5/36 (13.9%)	13/36 (36.1%)	10/36 (27.8%)
Cyber/Technical Fraud	8/36 (22.2%)	5/36 (13.9%)	9/36 (25.0%)	7/36 (19.4%)
Social Relationship Scams	10/36 (27.8%)	7/36 (19.4%)	8/36 (22.2%)	6/36 (16.7%)
Consumer Fraud	13/36 (36.1%)	10/36 (27.8%)	7/36 (19.4%)	5/36 (13.9%)
Government/Institutional Impersonation Fraud	8/36 (22.2%)	6/36 (16.7%)	14/36 (38.9%)	10/36 (27.8%)
Social Media Scams	14/36 (38.9%)	11/36 (30.6%)	7/36 (19.4%)	5/36 (13.9%)
Total	**60/216 (27.8%)**	**44/216 (20.4%)**	**58/216 (26.9%)**	**43/216 (19.9%)**

Note. Values are presented as deceptive or risky responses/total exposures (%). A deceptive or risky response refers to a judgment or intended action indicating vulnerability to the fraud scenario, such as misclassifying the message as safe or expressing willingness to comply with the fraudulent request. YA = younger group; OA = older group.

Table 12
Multilevel logistic regression results for deceptive-response likelihood in RQ2 with effect sizes.

Predictor	Estimate	SE	Test statistic	p	Effect size
AI condition (AI vs. control)	−0.42	0.13	$z = -3.23$	0.001	OR = 0.66
Age group (older vs. younger)	−0.04	0.12	$z = -0.33$	0.734	OR = 0.96
Fraud type	–	–	$\chi^2(5) = 24.81$	< 0.001	$w = 0.17$
AI × Age group	0.03	0.10	$z = 0.30$	0.781	OR = 1.03
Age group × Fraud type	–	–	$\chi^2(5) = 17.66$	0.003	$w = 0.14$
AI × Fraud type	–	–	$\chi^2(5) = 10.94$	0.052	$w = 0.11$
AI × Age group × Fraud type	–	–	$\chi^2(5) = 9.87$	0.079	$w = 0.11$

Note. The dependent variable was deceptive response likelihood at the scenario level. Negative coefficients for AI condition indicate reduced susceptibility in the AI-assisted condition. Odds ratios (ORs) are reported as effect sizes for model coefficients, and Cohen's w is reported for omnibus χ^2 effects. **Practical significance:** AI support was associated with approximately 34% lower odds of deceptive responses, while the small fraud-type and age-by-fraud-type effects indicate category-specific vulnerability patterns; the AI-by-fraud-type and three-way effects should be interpreted as descriptive trends.

Table 13
Reduction in deceptive responses from control to AI-assisted condition by age group and fraud type.

Fraud type	Younger group: Absolute reduction	Younger group: Relative reduction	Older group: Absolute reduction	Older group: Relative reduction
Financial/Payment Fraud	5.5%	28.6%	8.3%	23.1%
Cyber/Technical Fraud	8.3%	37.5%	5.6%	22.2%
Social Relationship Scams	8.4%	30.0%	5.5%	25.0%
Consumer Fraud	8.3%	23.1%	5.5%	28.6%
Government/Institutional Impersonation Fraud	5.5%	25.0%	11.1%	28.6%
Social Media Scams	8.3%	21.4%	5.5%	28.6%

Note. Absolute reduction refers to the percentage-point decrease in deceptive responses between the control and AI-assisted conditions. Relative reduction refers to the proportional decrease relative to the control-group baseline within the same age group.

condition, these rates declined to 30.6% and 27.8%, respectively, indicating meaningful reductions in the scam categories to which younger participants were initially most vulnerable. Among older group, the highest deceptive-response rates in the control condition were observed for *government or institutional impersonation fraud* (38.9%) and *financial/payment fraud* (36.1%). In the AI-assisted condition, both rates declined to 27.8%. These descriptive patterns suggest that although vulnerability profiles differed across age groups, AI-assisted support was associated with lower susceptibility in the fraud categories that posed the greatest risk to each group.

For completeness, Table 13 reports the absolute and relative reduction in deceptive responses for each fraud category within each age group. The largest reductions for the younger group were observed in *social media scams*, followed by *consumer fraud*, whereas the largest reductions for the older group were observed in *government or institutional impersonation fraud* and *financial/payment fraud*. This age-specific pattern is consistent with the theoretical expectation that fraud vulnerability is shaped by an interaction between user characteristics and task characteristics, and that AI support is especially useful when it helps users identify the deceptive cues most relevant to the scam types to which they are most vulnerable.

Findings for RQ2: Fraud vulnerability followed different scam-type patterns in younger and older groups, rather than being uniformly distributed across age groups. Importantly, AI-assisted decision support substantially reduced deceptive responses in the scam categories that posed the greatest risk for each group, suggesting that the intervention was effective in addressing age-specific vulnerability profiles.

4.4. Participants' perceptions and experiences of the AI-assisted decision support system (RQ3)

In this section, we examined participants' post-study evaluations of the AI-assisted decision support system and analyzed the qualitative interview data collected after the 10-day exposure period. This part of the analysis focused on how participants perceived the system in terms of *usefulness*, *trust*, *realism*, and *confidence*, as well as how they described their actual experiences of using the AI support while responding to suspicious messages in ecologically valid fraud scenarios.

Table 14 presents the descriptive results for the post-study evaluative measures in the AI-assisted condition by age group. Overall, participants in both age groups reported relatively positive evaluations of the intervention. Perceived usefulness received the highest ratings in both groups, suggesting that participants generally viewed the AI system as helpful in making safer fraud-related decisions. Trust in AI-generated advice was also moderately high, although younger group

Table 14
Post-study evaluative measures in the AI-assisted condition by age group with effect sizes.

Measure	Younger group	Older group	t/Test statistic	p	Effect size
Perceived usefulness of the AI system, Mean (SD)	4.08 (0.52)	4.24 (0.49)	$t(34) = -0.95$	0.349	$d = -0.32$
Trust in AI-generated advice, Mean (SD)	3.91 (0.56)	3.68 (0.61)	$t(34) = 1.18$	0.246	$d = 0.39$
Perceived realism of the fraud scenarios, Mean (SD)	4.02 (0.47)	4.11 (0.50)	$t(34) = -0.56$	0.579	$d = -0.19$
Self-reported confidence in fraud detection after the study, Mean (SD)	3.84 (0.51)	4.03 (0.55)	$t(34) = -1.08$	0.287	$d = -0.36$
Intention to use a similar AI tool in daily life, Mean (SD)	4.16 (0.58)	3.89 (0.63)	$t(34) = 1.34$	0.189	$d = 0.45$

Note. Ratings were recorded on a 5-point Likert scale, with higher scores indicating more positive evaluations. Results are reported for participants in the AI-assisted condition only. Cohen's d is reported as the effect size for age-group comparisons, with positive values indicating higher ratings in the younger group. Practical significance: between-age differences were small to moderate, suggesting that both age groups evaluated the AI-assisted system broadly positively.

Table 15
One-sample tests of post-study evaluative measures against the scale midpoint with effect sizes.

Measure	Group	Mean (SD)	t	p	Effect size
Perceived usefulness of the AI system	Younger group	4.08 (0.52)	8.78	< 0.001	$d_z = 2.07$
	Older group	4.24 (0.49)	10.72	< 0.001	$d_z = 2.53$
Trust in AI-generated advice	Younger group	3.91 (0.56)	6.88	< 0.001	$d_z = 1.62$
	Older group	3.68 (0.61)	4.76	< 0.001	$d_z = 1.12$
Perceived realism of the fraud scenarios	Younger group	4.02 (0.47)	9.10	< 0.001	$d_z = 2.14$
	Older group	4.11 (0.50)	9.39	< 0.001	$d_z = 2.21$
Self-reported confidence in fraud detection after the study	Younger group	3.84 (0.51)	6.99	< 0.001	$d_z = 1.65$
	Older group	4.03 (0.55)	7.95	< 0.001	$d_z = 1.87$
Intention to use a similar AI tool in daily life	Younger group	4.16 (0.58)	8.53	< 0.001	$d_z = 2.01$
	Older group	3.89 (0.63)	5.35	< 0.001	$d_z = 1.26$

Note. The comparison midpoint was 3 on a 5-point Likert scale. Cohen's d_z is reported as the effect size for one-sample comparisons. Practical significance: the large effect sizes indicate that both age groups rated the AI-assisted system meaningfully above the neutral midpoint across usefulness, trust, realism, confidence, and future-use intention.

reported slightly higher trust than older group. At the same time, older group gave somewhat higher ratings to the perceived usefulness of the system and to the confidence they felt after using it, indicating that although they were initially more cautious, they may have benefited more strongly from concrete decision support once they engaged with it. Ratings of scenario realism were high in both groups, suggesting that the fraud materials were experienced as sufficiently plausible and relevant to everyday digital communication.

To further examine participants' overall attitudes toward the intervention, we conducted one-sample comparisons against the scale midpoint. As shown in Table 15, perceived usefulness, perceived realism, and intention to use a similar AI tool in daily life were significantly above the midpoint in both age groups. Trust in AI-generated advice was also above the midpoint, although the effect was somewhat weaker among the older group. These results suggest that participants generally viewed the AI-supported intervention as meaningful, usable, and relevant, even if their degree of trust varied somewhat across age groups.

Taken together, the quantitative evaluative measures indicate that the AI-assisted system was generally well received. Participants in both age groups considered it useful and realistic, and most indicated that they would be willing to use a similar tool in daily life. At the same time, the pattern of means suggests a subtle age-related difference in how participants approached the intervention: the younger group reported somewhat greater trust in the AI output and stronger willingness to use such a system again, whereas the older group reported slightly higher usefulness and greater post-study confidence. This pattern is consistent with the possibility that the younger group was more comfortable adopting the tool, while the older group may have depended more strongly on the guidance once they decided to use it.

The interview data provided further insight into how participants experienced the AI-assisted system and how they interpreted fraud-related decision making in realistic digital contexts. Thematic analysis identified four recurring themes: *AI as immediate support under uncertainty*, *trust shaped by explanation rather than automation alone*, *low-cost scams as psychologically easy to accept*, and *age-specific interpretations of credibility and compliance*. Table 16 summarizes these themes and their implications for understanding participants' experiences of the intervention.

Participants across both age groups frequently described the AI system as particularly valuable when suspicious messages arrived at moments of uncertainty and immediate human advice was not available. As one younger participant noted, *"When these messages come suddenly, you usually don't stop to think carefully. If the AI immediately points out the risky part, it feels like someone is reminding you to slow down"*. A similar pattern was reflected in the responses of other younger participants, many of whom described the system as a form of instant second opinion that was especially useful during ambiguous or time-pressured situations. Older participants also emphasized this immediacy, although they were more likely to describe the system in practical rather than technological terms. One older participant explained, *"If nobody around me answers in time, having something that tells me what to pay attention to is still better than deciding by myself in a hurry"*.

At the same time, the interviews showed that trust in the AI system was not unconditional. Participants repeatedly stated that they were more willing to rely on the AI output when it included an explanation of suspicious cues rather than only a risk label. One younger participant commented, *"I don't just want it to say high risk. I want to know why. If it tells me the link is strange or the message is using pressure, then I trust*

Table 16
Main themes from the post-study interviews.

Theme	Description	Interpretive significance
AI as immediate support under uncertainty	Participants described the AI system as useful when suspicious messages arrived unexpectedly and there was no one immediately available to consult.	Highlights the practical value of AI at the moment of risk, especially in time-sensitive fraud situations.
Trust shaped by explanation rather than automation alone	Participants were more willing to rely on AI when the system explained *why* a message appeared risky rather than simply labeling it as dangerous.	Suggests that interpretability is central to the acceptance of AI-supported fraud detection.
Low-cost scams as psychologically easy to accept	Participants reported that small payments, deposits, or minor fees often felt less threatening and therefore more "worth trying."	Explains why some fraud scenarios elicited deceptive responses despite awareness of possible risk.
Age-specific interpretations of credibility and compliance	Younger and older groups described different reasons for trusting suspicious messages, including platform familiarity, social convenience, authority cues, and politeness.	Supports the finding that vulnerability differs across age groups not only in degree, but also in underlying decision logic.

it more". Similar comments were made by other younger participants, who often framed trust in terms of transparency and logic. Among the older group, trust was more mixed. Some participants appreciated the clarity of the recommended action, but others expressed hesitation toward relying on an electronic system. One older participant stated, *"A machine has no feelings. Sometimes I still don't feel comfortable trusting it completely, even if it seems useful"*. Several other older participants voiced comparable reservations, suggesting that emotional distance from digital devices may be one source of resistance to AI-supported fraud prevention.

The interviews also helped explain why certain scam categories were especially persuasive. Across both age groups, many deceptive responses were associated with scams involving *small monetary amounts* or requests that appeared minor, routine, or easy to reverse. Younger participants frequently described these scenarios as not seeming serious enough to justify strong caution. One younger participant remarked, *"If it's only a small amount, like a deposit or a fee to join something, people think maybe it's worth trying"*. Similar statements were made by other younger participants, especially in relation to social media and consumer fraud scenarios. In these cases, the low-cost framing appeared to reduce perceived risk and encourage exploratory compliance. This pattern helps explain why the younger group was particularly vulnerable to promotion-based or platform-based scams involving apparently manageable amounts of money.

For older group, by contrast, the interviews suggested that trust was more often shaped by cues of authority, politeness, and procedural legitimacy. Several older participants reported difficulty distinguishing genuine from fake government or institutional representatives, especially when the communication was delivered in a calm, formal, or respectful manner. One older participant explained, *"If someone says they are from a government office and they speak politely and clearly, it is easy to think they are genuine"*. Other older participants described similar experiences, noting that they were more likely to trust authority-based requests when they appeared orderly, routine, and not excessively aggressive. This pattern was especially evident in the government impersonation and financial/payment scenarios, where small fees or official-sounding requests were more readily interpreted as plausible.

Overall, the qualitative findings complement the quantitative results by showing that participants generally perceived the AI-assisted system as useful and realistic, but that their trust in and reliance on the system depended on both the form of the AI explanation and the nature of the fraud context. Younger participants tended to value speed, flexibility, and cue-based reasoning, while older participants relied more heavily on concrete recommendations once they engaged with the system, even though some remained hesitant toward technology itself. The interview data therefore reinforce the conclusion that the effectiveness of AI-assisted fraud detection is shaped not only by the presence of the tool, but also by how different users interpret its role in relation to uncertainty, credibility, and everyday decision-making.

Findings for RQ3: Participants in both age groups evaluated the AI-assisted system positively in terms of usefulness, realism, and post-study confidence, while trust in the system depended strongly on whether the AI explanation was interpretable and actionable. The interview data further showed that younger and older groups differed in how they experienced scam credibility and AI support.

5. Discussion

5.1. Principal findings

The present study investigated whether AI-assisted decision support can reduce susceptibility to digital fraud in ecologically valid settings and whether its effects vary across age groups and fraud categories. Taken together, the findings across RQ1–RQ3 provide a coherent picture. First, AI-assisted support significantly improved fraud detection accuracy and reduced risky behavioral intention relative to unaided human judgment. This finding extends prior fraud research, which has largely focused on identifying risk factors associated with victimization rather than testing real-time intervention mechanisms in realistic decision contexts (Dadà et al., 2025; Norris et al., 2019). In this sense, the present study provides evidence that timely, task-sensitive support can reduce fraud vulnerability at the point of risk, when suspicious messages are encountered and users must decide how to respond. This result is consistent with prior evidence that forewarning can reduce fraud susceptibility among vulnerable consumers (Scheibe et al., 2014). However, the present study extends this prevention logic by embedding warnings within a real-time AI-assisted decision support system that also explained suspicious cues and recommended safer actions. The decision-time findings further suggest that AI support may have changed not only whether participants identified fraud correctly, but also how they approached the decision. Participants in the AI-assisted condition took longer to complete each scenario than those in the control condition. Because this increase in decision time occurred alongside higher detection accuracy and lower risky behavioral intention, it may reflect more deliberative processing rather than poorer task performance. This interpretation is consistent with the design of the AI support, which encouraged participants to review risk assessments, suspicious cues, and recommended actions before submitting their final responses.

Second, our results reinforce the view that fraud vulnerability is *patterned rather than uniform*. Consistent with the systematic review by Dadà et al. (2025), the younger and older groups did not display a single, stable hierarchy of vulnerability; rather, their susceptibility varied across fraud categories. In our study, the younger group appeared more vulnerable to fraud scenarios embedded in social-media-like and consumer-oriented contexts, whereas the older group appeared more vulnerable to authority-based and financially framed scenarios, particularly government/institutional impersonation fraud and financial/payment fraud. This pattern is broadly consistent with prior work

suggesting that younger people may be more exposed to and involved in online and platform-based fraud contexts, whereas older individuals may face higher victimization risk in certain financially consequential or institutionally framed fraud contexts (DeLiema et al., 2024; Kemp & Erades Pérez, 2023; Xu et al., 2022; Yu et al., 2023). From the perspective of the Person–Task Fit framework, these differences suggest that fraud vulnerability may emerge when the demands of a deceptive task align with, or exceed, users' expectations, routines, and available evaluative resources. Social-media and consumer-oriented scenarios may have appeared more familiar, routine, or low-risk to the younger group, whereas financial pressure and institutional impersonation may have carried stronger credibility cues for the older group. Our findings therefore support the argument that age differences in fraud are better understood as *fraud-category-specific vulnerability profiles* than as a simple overall age effect. Although the interaction terms involving AI condition and fraud category should be interpreted cautiously, the descriptive reductions suggest that AI-assisted support may be particularly useful when it highlights risk cues that users might otherwise underestimate.

Third, the present findings highlight the importance of interpreting fraud as an active decision-making process rather than merely a background trait or static victim profile. This aligns closely with the review's use of the Person–Task Fit perspective, according to which fraud vulnerability emerges from the interaction between decision-maker characteristics, scam characteristics, and contextual conditions (Dadà et al., 2025). In our data, this interaction was evident not only in the age-group differences across scam types, but also in the way participants responded to low-cost and apparently manageable fraudulent requests. As the interview data suggested, many deceptive responses were associated with scams framed as small, routine, or socially convenient. This interpretation resonates with prior research showing that scam susceptibility is shaped not only by demographic factors, but also by impulsivity, confidence, digital literacy, financial judgment, and the structure of the decision environment itself (Du & Chen, 2023; Kirwan et al., 2018; Sarno & Black, 2024).

Fourth, the AI-assisted intervention appears to have been especially valuable in the scam categories that posed the greatest baseline risk for each age group. This is an important contribution because the broader literature has repeatedly noted that while many studies identify who is vulnerable, much less is known about which interventions actually reduce deception at the point of exposure (Norris et al., 2019). In our study, AI support did not merely improve overall performance; it helped lower deceptive responses in the precise scam categories to which each group was most vulnerable. This pattern suggests that AI-based support may be most effective when it helps users interpret the cues that are especially difficult for them in a given fraud context. In this respect, our findings complement prior phishing and online-fraud studies showing that suspicion, discrimination accuracy, cognitive reflection, and decision style play important roles in detecting deceptive communications (Butavicius et al., 2022; Jones et al., 2019; Sarno & Black, 2024).

Finally, the qualitative findings add an important layer to the interpretation of these quantitative results. Participants did not respond to the AI system in a purely technical manner; instead, they evaluated it through the lens of trust, immediacy, and interpretability. Many participants appreciated the system because it provided immediate support when human advice was unavailable, but they were more willing to rely on it when the output explained *why* a message was suspicious rather than merely labeling it as high risk. This finding is especially important in light of prior work showing that fraud decisions are often made under uncertainty, time pressure, and incomplete verification opportunities (Dadà et al., 2025; Grilli et al., 2021; Sarno et al., 2020). It also helps explain why the older group in our study relied more heavily on concrete recommended actions once they engaged with the AI system, even though some remained initially cautious toward digital tools.

Taken together, the present study offers evidence that AI-assisted decision support can play a meaningful role in reducing digital fraud susceptibility. The findings suggest that fraud prevention is most effective when it is immediate, interpretable, and sensitive to the ways in which different users experience different scam types. Rather than assuming that some people are simply resistant to fraud while others are not, the results point toward a more realistic conclusion: vulnerability depends on the fit between the person and the scam. In this sense, the contribution of AI is not merely to detect fraud on behalf of users, but to support safer judgment when people encounter precisely the kind of deception that they are most likely to underestimate.

5.2. Implications

The results carry several practical implications. First, AI-supported fraud prevention tools should be designed as *interpretable decision aids* rather than as opaque warning systems. Risk labels alone may not be sufficient; users appear more likely to trust and use AI support when it highlights specific suspicious cues and offers clear recommended actions. Second, intervention design should be tailored to scam type and user profile. For younger users, it may be especially important to emphasize the hidden risk of low-cost, socially embedded, or promotional scams that appear minor or reversible. For older users, greater attention may be needed for authority-based and institutionally framed deception, especially where politeness, procedural formality, and small official-looking fees create a false sense of legitimacy. Third, the strong role of low-cost fraudulent requests in the interview data suggests that anti-fraud interventions should not focus exclusively on large or obviously dangerous losses. Small payments, deposits, and low-threshold requests may be especially powerful precisely because they appear too minor to trigger careful scrutiny.

These findings also align with recent prevention research suggesting that the delivery format of anti-fraud advice matters. Broad awareness campaigns or generic mass messages may have limited effects when users encounter fraud under time pressure or emotional arousal (Jensen et al., 2025; Kircanski et al., 2018). For older or more isolated users, fraud-prevention advice may be more effective when it is delivered through trusted channels and supported by interpersonal or one-to-one guidance (Button et al., 2024). AI-assisted decision support may therefore be most useful when it functions not merely as an automated warning label, but as an immediate, interpretable, and action-oriented source of support that helps users verify suspicious requests at the moment of exposure.

Several broader recommendations follow from these findings. Educational programs should train users not only to recognize obviously suspicious messages, but also to question requests that appear *small*, *routine*, or *socially convenient*. Public communication campaigns may benefit from emphasizing that serious fraud does not always begin with a large transfer or dramatic threat; it often begins with a minor action that feels harmless. Similarly, designers of digital platforms and messaging systems should consider integrating just-in-time warning mechanisms that make deceptive cues visible before users take action. In institutional settings, especially those involving older participants, AI-supported fraud warnings may be most effective when paired with simple behavioral recommendations, such as verifying through official channels, delaying payment, or consulting a trusted contact.

The findings also have implications for the channel design of AI-assisted fraud prevention. Although the present study focused on web-based simulated digital messages, fraud exposure is not limited to text-based online communication. Recent research on telephone fraud among older adults shows that many attempted frauds still occur through phone-based channels, particularly among older and more socially isolated groups (Button et al., 2025). In many real-world cases, an initial digital message may function only as the "hook", after which the more intensive social engineering occurs through telephone calls, voice messages, or other interpersonal communication channels.

This suggests that AI-assisted fraud prevention should not be limited to detecting suspicious written messages. Future systems may need to support cross-channel risk assessment, including text, links, phone numbers, call context, voice-based warnings, and guidance for verifying suspicious requests through trusted channels.

5.3. Limitations and suggestions

The present study should be interpreted in light of several limitations. First, the sample was designed as a balanced online lab-in-the-field experimental sample across age groups and AI conditions rather than as a demographically representative sample of the wider population. The total sample size was modest, the younger group included a relatively high proportion of students, and the older group covered a broad age range of 50 years and above. In addition, because eligibility required regular use of a smartphone or computer, individuals with limited digital engagement may be underrepresented, including some older adults who may be more exposed to telephone-based fraud or rely less frequently on online communication technologies. This is particularly relevant because telephone-based scams remain important among older populations, while social isolation and living arrangements may further influence fraud vulnerability and access to timely support (Button et al., 2025). The modest participant sample also limits statistical sensitivity. The sensitivity analysis indicated that, under a conservative between-participant specification with $\alpha = 0.05$ and power of 0.80, the study could reliably detect effects of approximately $f = 0.335$ or larger. Although the repeated-exposure design provided multiple scenario-level observations that were appropriately modeled using multilevel analyses, these observations do not substitute for a larger number of independent participants; consequently, smaller effects and higher-order interactions involving AI condition, age group, and fraud type may have been more difficult to detect reliably. The findings should therefore be interpreted as evidence from two age-defined groups with regular digital engagement rather than as population-level estimates, and non-significant higher-order interactions should be interpreted cautiously. Future studies should employ larger and more demographically representative samples, distinguish more fine-grained age groups, and extend AI-assisted fraud support to telephone-based, voice-based, and cross-channel fraud scenarios.

Second, although the fraud scenarios were designed to be ecologically realistic, they remained simulated rather than genuinely consequential. Participants encountered suspicious communications in contexts intended to resemble everyday digital fraud exposure, but they did not face the full emotional, financial, and social consequences associated with real victimization. In actual fraud situations, stress, embarrassment, urgency, social pressure, and the perceived irreversibility of a decision may operate more strongly than in experimental contexts. Accordingly, the present findings should be interpreted as evidence of behavior under conditions designed to enhance ecological realism, but not as a complete substitute for real-world fraud outcomes. Furthermore, because the study was conducted in mainland China and Hong Kong, the findings should be interpreted in relation to these socio-technical and regulatory contexts. Fraud categories, communication channels, platform use, reporting systems, and trust in institutions may vary across countries and regions. Future research should examine whether similar AI-assisted decision support effects are observed in other national contexts and should adapt scenario categories to locally prevalent fraud typologies.

Last but not least, the AI-assisted framework was evaluated as a prototype decision-support system rather than as a fully mature or widely deployed intervention tool. Its effectiveness may therefore depend on specific design choices, including the clarity of its explanations, the format of its warnings, the timing of its support, and the extent to which users actively engage with the output. In particular, the interview findings suggest that participants did not respond only to the presence of AI support, but also to how interpretable, concrete, and actionable that support appeared. Future studies should therefore compare alternative forms of AI explanation, different interface designs, and different levels of automation in order to determine which configurations are most effective for fraud prevention.

6. Conclusion

This study examined whether AI-assisted decision support can help reduce susceptibility to digital fraud in ecologically valid settings and whether its effects vary across age groups and fraud categories. Within this 10-day online lab-in-the-field experiment, AI-assisted support was associated with higher fraud detection accuracy, lower risky behavioral intention, and safer intended responses when participants encountered simulated suspicious communications. At the same time, the younger and older groups showed different fraud-category-specific vulnerability patterns, with the younger group showing greater susceptibility to social-media- and consumer-oriented scenarios and the older group showing greater susceptibility to authority-based and financially framed scenarios. Post-study questionnaire ratings and interview findings further indicated that participants generally perceived the AI system as useful and realistic, while trust in the system appeared to depend on whether its guidance was interpretable, concrete, and actionable. Overall, the study suggests that digital fraud prevention may benefit from moving beyond static assumptions about who is or is not vulnerable and from exploring timely, explainable, and context-sensitive AI support as one possible way to help users respond more safely to suspicious communications.

CRediT authorship contribution statement

Yicheng Sun: Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jacky Keung:** Supervision, Software, Resources, Methodology, Funding acquisition, Formal analysis. **Hi Kuen Yu:** Writing – review & editing, Validation, Resources, Data curation, Conceptualization. **Yuchen Cao:** Validation, Methodology, Investigation.

Ethics declaration

All procedures involving human participants were conducted in accordance with the ethical principles of the Declaration of Helsinki. The study involved simulated digital fraud scenarios only and did not involve real financial transactions, collection of genuine personal credentials, or interaction with actual fraudulent websites or external links. Participants were informed of the study purpose, procedures, data collection practices, voluntary nature of participation, and their right to withdraw before providing informed consent. All collected data were anonymized and used solely for research purposes.

Declaration of AI Tools Used

An AI-assisted decision support prototype was used as the intervention tool in this study. For participants assigned to the AI-assisted condition, the system provided fraud risk assessments, explanations of suspicious cues, and recommended actions for simulated digital fraud scenarios. The AI tool was used only within the controlled experimental tasks and did not make decisions on behalf of participants or process real financial credentials. All materials, AI-supported outputs, analyses, and interpretations were reviewed and verified by the authors, who take full responsibility for the final manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Baseline questionnaire

This questionnaire was administered before the 10-day fraud-exposure phase. All items were structured quantitative items. Unless otherwise stated, Likert-scale items were rated from 1 = strongly disagree to 5 = strongly agree.

A. Demographic and background information

1. Age: _____
2. Gender:
   (a) Male
   (b) Female
   (c) Prefer not to say
3. Employment status:
   (a) Student
   (b) Employed full-time
   (c) Employed part-time
   (d) Retired
   (e) Other
4. Primary device used in daily life:
   (a) Smartphone mainly
   (b) Computer mainly
   (c) Both equally
5. Average daily digital use:
   (a) Less than 2 h
   (b) 2–4 h
   (c) More than 4 h

B. Digital literacy

Please indicate your level of agreement with each statement.

6. I am familiar with common digital communication tools, such as email, text messaging, social media, and mobile applications.
7. I can usually identify suspicious links in emails, text messages, or social media messages.
8. I know how to check whether an online message or website is authentic.
9. I am confident in using official websites or apps to verify suspicious information.
10. I understand the basic risks of sharing personal or financial information online.

C. Prior fraud exposure and scam-related harm

11. Have you previously encountered suspicious or potentially fraudulent messages, such as phishing emails, suspicious payment requests, impersonation attempts, fraudulent links, or scam-related messages?
    (a) Yes
    (b) No
12. Through which channels have you encountered suspicious or potentially fraudulent messages? Select all that apply.
    (a) Email
    (b) Text message or SMS
    (c) Phone call
    (d) Social media platform
    (e) Mobile application notification
    (f) Online shopping or delivery platform
    (g) I have not encountered such messages
13. Have you previously been personally affected by a scam, such as through financial loss, disclosure of personal information, clicking a fraudulent link, or taking another action that led to a negative consequence?
    (a) Yes
    (b) No
14. If yes, what type of consequence did you experience? Select all that apply.
    (a) Financial loss
    (b) Disclosure of personal information
    (c) Disclosure of account or login information
    (d) Clicking a fraudulent link
    (e) Continuing communication with a fraudulent sender
    (f) Other negative consequence
    (g) Not applicable

D. Risk perception

Please indicate your level of agreement with each statement.

15. Digital communication can involve serious fraud risks.
16. I may be vulnerable to online scams in everyday digital communication.
17. I believe digital fraud is a realistic threat in daily life.
18. I usually treat unexpected requests for money, personal information, or account verification as potentially risky.

E. Confidence in detecting suspicious messages

Please indicate your level of agreement with each statement.

19. I am confident that I can distinguish legitimate messages from fraudulent messages.
20. I am confident that I can decide whether a suspicious message requires further verification.
21. I am confident that I can avoid unsafe actions, such as clicking suspicious links or providing sensitive information.

Scoring and use in analysis

Digital literacy was calculated as the mean score of Items 6–10. Risk perception was calculated as the mean score of Items 15–18. Confidence in detecting suspicious messages was calculated as the mean score of Items 19–21. Prior fraud exposure was coded from Item 11 as 1 = yes and 0 = no. Previous scam-related harm was coded from Item 13 as 1 = yes and 0 = no. Background variables were summarized descriptively and used to assess group comparability before the intervention.

Data availability

Data will be made available on request. Requests for access should be directed to Mr. Yicheng Sun, subject to ethical approval and data protection requirements.

References

Anti-Deception Coordination Centre (2026). Scam situation in Hong Kong. https://www.adcc.gov.hk/en-hk/statistic.html.

Baki, S., & Verma, R. M. (2022). Sixteen years of phishing user studies: What have we learned? *IEEE Transactions on Dependable and Secure Computing*, *20*(2), 1200–1212.

Bar Lev, E., Maha, L. G., & Topliceanu, S. C. (2022). Financial frauds' victim profiles in developing countries. *Frontiers in Psychology*, *13*, Article 999053.

Bele, J. L. (2020). Financial scams, frauds, and threats in the digital age. *Modern Approaches To Knowledge Management Development*, 39.

Burton, A., Cooper, C., Dar, A., Mathews, L., & Tripathi, K. (2022). Exploring how, why and in what contexts older adults are at risk of financial cybercrime victimisation: A realist review. *Experimental Gerontology*, *159*, Article 111678.

Butavicius, M., Taib, R., & Han, S. J. (2022). Why people keep falling for phishing scams: The effects of time pressure and deception cues on the detection of phishing emails. *Computers & Security*, *123*, Article 102937.

Button, M., Shepherd, D., Hawkins, C., & Tapley, J. (2024). Disseminating fraud awareness and prevention advice to older adults: perspectives on the most effective means of delivery. *Crime Prevention and Community Safety*, *26*(4), 385–400.

Button, M., Shepherd, D., Hawkins, C., & Tapley, J. (2025). Fear and phoning: Telephones, fraud, and older adults in the UK. *International Review of Victimology*, *31*(1), 117–134.

Charness, G., Gneezy, U., & Kuhn, M. A. (2013). Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics. *Journal of Economic Behavior and Organization*, *91*, 93–100.

Chen, Y., & Konstan, J. (2015). Online field experiments: a selective survey of methods. *Journal of the Economic Science Association*, *1*(1), 29–42.

Colautti, L., Robba, M., Antonietti, A., & Iannello, P. (2025). Profiling older adults' decision-making under risk: the role of cognitive functioning and personality traits. *Thinking & Reasoning*, *31*(4), 474–496.

Dadà, C. B., Colautti, L., Rosi, A., Cavallini, E., Antonietti, A., & Iannello, P. (2025). Uncovering vulnerability to fraud and scams among adult victims in online and offline contexts: A systematic review. *Computers in Human Behavior*, *172*, Article 108734.

DeLiema, M., Langton, L., Brannock, D., & Preble, E. (2024). Fraud victimization across the lifespan: evidence on repeat victimization using perpetrator data. *Journal of Elder Abuse & Neglect*, *36*(3), 227–250.

DeLiema, M., Li, Y., & Mottola, G. (2023). Correlates of responding to and becoming victimized by fraud: Examining risk factors by scam type. *International Journal of Consumer Studies*, *47*(3), 1042–1059.

Du, W., & Chen, M. (2023). Too much or less? The effect of financial literacy on resident fraud victimization. *Computers in Human Behavior*, *148*, Article 107914.

Dupuis, D., Smith, D., & Gleason, K. (2023). Old frauds with a new sauce: digital assets and space transition. *Journal of Financial Crime*, *30*(1), 205–220.

Fan, J. X., & Yu, Z. (2022). Prevalence and risk factors of consumer financial fraud in China. *Journal of Family and Economic Issues*, *43*(2), 384–396.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Federal Trade Commission (2025). New FTC data show a big jump in reported losses to fraud to $12.5 billion in 2024. https://www.ftc.gov/news-events/news/press-releases/2025/03/new-ftc-data-show-big-jump-reported-losses-fraud-125-billion-2024.

Ferreira, A., Coventry, L., & Lenzini, G. (2015). Principles of persuasion in social engineering and their use in phishing. In *International conference on human aspects of information security, privacy, and trust* (pp. 36–47). Springer.

Ferreira, A., & Teles, S. (2019). Persuasion: How phishing emails can influence users and bypass security measures. *International Journal of Human-Computer Studies*, *125*, 19–31.

Gangadharan, L., Jain, T., Maitra, P., & Vecci, J. (2022). Lab-in-the-field experiments: perspectives from research on gender. *The Japanese Economic Review*, *73*(1), 31–59.

Grilli, M. D., McVeigh, K. S., Hakim, Z. M., Wank, A. A., Getz, S. J., Levin, B. E., Ebner, N. C., & Wilson, R. C. (2021). Is this phishing? Older age is associated with greater difficulty discriminating between safe and malicious emails. *The Journals of Gerontology: Series B*, *76*(9), 1711–1715.

Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*(4), 1009–1055.

Hong Kong Police Force (2026). Law and order situation in Hong Kong in 2025. https://www.police.gov.hk/ppp_en/01_about_us/cp_ye.html.

Irvin-Erickson, Y. (2024). Identity fraud victimization: A critical review of the literature of the past two decades. *Crime Science*, *13*(1), 3.

Jair, H., & Mustafa, K. (2025). The rise of economic crime in the digital age: Cryptocurrency fraud and online scams.

James, B. D., Boyle, P. A., & Bennett, D. A. (2014). Correlates of susceptibility to scams in older adults without dementia. *Journal of Elder Abuse & Neglect*, *26*(2), 107–122.

Jensen, R. I. T., Gerlings, J., & Ferwerda, J. (2025). Do awareness campaigns reduce financial fraud? RIT jensen et al.. *European Journal on Criminal Policy and Research*, *31*(4), 717–752.

Jones, H. S., Towse, J. N., Race, N., & Harrison, T. (2019). Email fraud: The search for psychological predictors of susceptibility. *PLoS One*, *14*(1), Article e0209684.

Junger, M., Koning, L., Hartel, P., & Veldkamp, B. (2023). In their own words: deception detection by victims and near victims of fraud. *Frontiers in Psychology*, *14*, Article 1135369.

Kadoya, Y., Khan, M. S. R., Narumoto, J., & Watanabe, S. (2021). Who is next? A study on victims of financial fraud in Japan. *Frontiers in Psychology*, *12*, Article 649565.

Kemp, S., & Erades Pérez, N. (2023). Consumer fraud against older adults in digital society: Examining victimization and its impact. *International Journal of Environmental Research and Public Health*, *20*(7), 5404.

Kircanski, K., Notthoff, N., DeLiema, M., Samanez-Larkin, G. R., Shadel, D., Mottola, G., Carstensen, L. L., & Gotlib, I. H. (2018). Emotional arousal may increase susceptibility to fraud in older and younger adults.. *Psychology and Aging*, *33*(2), 325.

Kirwan, G. H., Fullwood, C., & Rooney, B. (2018). Risk factors for social networking site scam victimization among Malaysian students. *Cyberpsychology, Behavior, and Social Networking*, *21*(2), 123–128.

Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, *8*(1), 33267.

Ministry of Public Security of the People's Republic of China (2024). The ministry of public security announces ten high-frequency telecom and online fraud types. https://www.chinanews.com.cn/gn/2024/06-25/10240001.shtml.

National Financial Regulatory Administration (2024). Risk alert on preventing new forms of telecom and online fraud. https://www.nfra.gov.cn/cn/view/pages/ItemDetail.html?docId=1172651&itemId=4100.

Norris, G., Brookes, A., & Dowell, D. (2019). The psychology of internet fraud victimisation: A systematic review. *Journal of Police and Criminal Psychology*, *34*(3), 231–245.

Reshetnikova, N. N., Magomedov, M. M., Zmiyak, S. S., Gagarinskii, A. V., & Buklanov, D. A. (2021). Directions of digital financial technologies development: Challenges and threats to global financial security. In *Current problems and ways of industry development: equipment and technologies* (pp. 355–363). Springer.

Sarno, D. M., & Black, J. (2024). Who gets caught in the web of lies?: Understanding susceptibility to phishing emails, fake news headlines, and scam text messages. *Human Factors*, *66*(6), 1742–1753.

Sarno, D. M., Lewis, J. E., Bohil, C. J., & Neider, M. B. (2020). Which phish is on the hook? Phishing vulnerability for older versus younger adults. *Human Factors*, *62*(5), 704–717.

Scheibe, S., Notthoff, N., Menkin, J., Ross, L., Shadel, D., Deevy, M., & Carstensen, L. L. (2014). Forewarning reduces fraud susceptibility in vulnerable consumers. *Basic and Applied Social Psychology*, *36*(3), 272–279.

Shang, Y., Wu, Z., Du, X., Jiang, Y., Ma, B., & Chi, M. (2022). The psychology of the internet fraud victimization of older adults: A systematic review. *Frontiers in Psychology*, *13*, Article 912242.

State Council Information Office of the People's Republic of China (2026). China solves 258,000 telecom, online fraud cases in 2025. http://english.scio.gov.cn/pressroom/2026-01/08/content_118268937.html. (Accessed: 28 March 2026).

Ueno, D., Arakawa, M., Fujii, Y., Amano, S., Kato, Y., Matsuoka, T., & Narumoto, J. (2022). Psychosocial characteristics of victims of special fraud among Japanese older adults: A cross-sectional study using scam vulnerability scale. *Frontiers in Psychology*, *13*, Article 960442.

Wronska, M. K., Bujacz, A., Gocłowska, M. A., Rietzschel, E. F., & Nijstad, B. A. (2019). Person-task fit: Emotional consequences of performing divergent versus convergent thinking tasks depend on need for cognitive closure. *Personality and Individual Differences*, *142*, 172–178.

Xiang, H., Zhou, J., & Xie, B. (2020). Understanding older adults' vulnerability and reactions to telecommunication fraud: The effects of personality and cognition. In *International conference on human-computer interaction* (pp. 351–363). Springer.

Xu, L., Wang, J., Xu, D., & Xu, L. (2022). Integrating individual factors to construct recognition models of consumer fraud victimization. *International Journal of Environmental Research and Public Health*, *19*(1), 461.

Yu, L., Mottola, G., Kieffer, C. N., Mascio, R., Valdes, O., Bennett, D. A., & Boyle, P. A. (2023). Vulnerability of older adults to government impersonation scams. *JAMA Network Open*, *6*(9), Article e2335319.