

# SCVI: Bridging Social and Cyber Dimensions for Comprehensive Vulnerability Assessment

Shutonu Mitra  
*Virginia Tech*

Tomas Neguyen  
*Virginia Tech*

Qi Zhang  
*Virginia Tech*

Hyungmin Kim  
*Virginia Tech*

Hossein Salemi  
*George Mason University*

Chen-Wei Chang  
*Virginia Tech*

Fengxiu Zhang  
*George Mason University*

Michin Hong  
*Indiana University*

Chang-Tien Lu  
*Virginia Tech*

Hemant Purohit  
*George Mason University*

Jin-Hee Cho  
*Virginia Tech*

## Abstract

The rise of cyber threats on social media platforms necessitates advanced metrics to assess and mitigate social cyber vulnerabilities. This paper presents the Social Cyber Vulnerability Index (SCVI), a novel framework integrating individual-level factors (e.g., awareness, behavioral traits, psychological attributes) and attack-level characteristics (e.g., frequency, consequence, sophistication) for comprehensive socio-cyber vulnerability assessment. SCVI is validated using survey data (iPoll) and textual data (Reddit scam reports), demonstrating adaptability across modalities while revealing demographic disparities and regional vulnerabilities. Comparative analyses with the Common Vulnerability Scoring System (CVSS) and the Social Vulnerability Index (SVI) show SCVI's superior ability to capture nuanced socio-technical risks. Monte Carlo-based weight variability analysis confirms SCVI's robustness and highlights its utility in identifying high-risk groups. By addressing gaps in traditional metrics, SCVI offers actionable insights for policymakers and practitioners, advancing inclusive strategies to mitigate emerging threats such as AI-powered phishing and deepfake scams.

while the Social Vulnerability Index (SVI) identifies community vulnerabilities using socio-economic factors, it overlooks cyber-specific risks [15]. This work addresses these gaps by developing the Social Cyber Vulnerability Index (SCVI), a novel framework integrating individual-level characteristics and attack-specific threats, offering a specialized metric to inform policymakers and social media platforms about the vulnerabilities of high-risk demographics.

The importance of this work becomes evident in its effort to bridge the divide between social and technical vulnerabilities. SCVI leverages insights from studies highlighting the influence of behavioral, psychological, and sociodemographic factors on vulnerability [16, 27]. Unlike traditional metrics such as CVSS or the socio-technical Cyber Risk Index [7], SCVI emphasizes nuanced dimensions of human susceptibility, incorporating elements like psychological traits (e.g., impulsivity, credulity), past experiences with scams, and the sophistication of attack techniques. By integrating these diverse dimensions, SCVI enables a comprehensive evaluation of vulnerabilities across multiple contexts.

The methodological complexity of visualizing social cyber vulnerabilities further underscores the need for SCVI. Behavioral, social, and psychological factors are inherently multi-dimensional and context-dependent [25, 41]. Designing an index synthesizing these elements requires sophisticated analytical approaches, including sensitivity analyses and multi-dimensional visualizations. For instance, Alim et al. [3] proposed metrics such as Individual Vulnerability (VI) and Relative Vulnerability (VR) for social engineering risks but lacked validation and adaptability to real-world complexities. SCVI addresses these shortcomings, offering a robust and scalable framework validated through survey (iPoll) and social media (Reddit) datasets.

Finally, this work shows that user vulnerability can be assessed using survey responses and textual data from social media. Social media interactions provide insights into manipulative attempts, deceptive offers, and exploitation patterns targeting specific demographics [14, 35]. SCVI leverages these insights to highlight trends, enabling platforms to craft strate-

## 1 Introduction

### 1.1 Motivation & Goal

The increasing digital engagement of people on social media platforms has heightened their exposure to various cyber threats. Many individuals lack the digital literacy needed to recognize and mitigate these threats, making them prime targets for cybercriminals. The consequences of scams, phishing, and other cyberattacks are severe, resulting in substantial financial loss, emotional distress, and diminished trust in online services. Although the Common Vulnerability Scoring System (CVSS) remains a widely adopted standard for assessing technical vulnerabilities in systems and networks [6], no prior work has introduced a dedicated index to capture social cyber vulnerabilities affecting targeted populations. Similarly,

gies tailored to threats like AI-powered phishing or deepfake-related scams [29]. By encompassing diverse data sources, SCVI advances cybersecurity practices, ensuring inclusivity and adaptability in addressing evolving threats.

**The primary goal of this work** is to develop and validate SCVI, a comprehensive framework that integrates individual-level factors (e.g., awareness, psychological traits, behavioral characteristics) and attack-level characteristics (e.g., frequency, consequence, sophistication) [18, 30]. Through diverse data sources, this study aims to enhance socio-cyber vulnerability assessment, providing actionable insights for policymakers, practitioners, and platforms to design targeted interventions and close gaps in traditional metrics.

## 1.2 Key Contributions

This work makes the following **key contributions**:

1. **Innovative Social Cyber Vulnerability Metric:** This research introduces the Social Cyber Vulnerability Index (SCVI). This novel framework integrates individual-level factors (e.g., awareness, behavioral traits, psychological attributes) and attack-level characteristics (e.g., frequency, consequence, sophistication). SCVI addresses gaps in traditional metrics like CVSS and SVI, offering a comprehensive tool for assessing socio-cyber vulnerabilities. Its robust methodology, including feature extraction from survey and social media data and multi-dimensional vulnerability visualization, underscores its adaptability and utility.
2. **Validation Across Diverse Data Sources:** SCVI is validated using survey data (iPoll) and textual data (Reddit scam reports), demonstrating versatility across modalities. By incorporating individual and contextual factors, SCVI provides granular insights into cyber vulnerabilities, revealing geographical and demographic disparities that inform region-specific interventions and tailored strategies.
3. **Rigorous Empirical Validation and Comparison:** Sensitivity and weight variability analyses validate SCVI's reliability and robustness. Monte Carlo simulations highlight the dynamic contributions of individual and attack-level factors to SCVI variability. Comparative evaluations with CVSS and SVI confirm SCVI's superior ability to capture nuanced vulnerabilities while aligning with CVSS for technical risks and addressing gaps in socio-contextual indices like SVI.
4. **Practical Implications for Inclusive Cybersecurity:** SCVI supports targeted recommendations for policymakers, social media platforms, and cybersecurity practitioners by identifying high-risk groups and various scam typologies such as financial scams, phishing attacks, romance scams, online shopping fraud, tech support scams, etc. Its emphasis on demographic and sociocultural dimensions advances inclusivity. Future directions include expanding SCVI's geographic and demographic coverage, optimizing weighting with data-driven approaches, and adapting to

emerging threats like AI-powered phishing and deepfake fraud, ensuring its ongoing relevance.

## 2 Related Work

### 2.1 Cyber Vulnerability Metrics & Indices

Black et al. [6] evaluated CVSS for standardizing and assessing software vulnerabilities based on exploitability and impact on confidentiality, integrity, and availability. Flanagan et al. [15] introduced the Social Vulnerability Index (SVI) to identify vulnerable communities using socio-economic factors but did not address cyber vulnerabilities. Bolpagni [7] proposed a socio-technical Cyber Risk Index, linking higher HDI with lower cyber risk, though it excludes social media risks. Frauenstein and Flowerday [16] explored personality traits influencing phishing susceptibility but overlooked cultural factors and perceived risk. While CVSS and SVI overlook nuances in social engineering and modern cyber vulnerabilities, particularly in social media and evolving online threats, the Cyber Risk Index fails to address these threats' rapid evolution. Frauenstein and Flowerday [16] highlighted personality traits but neglected broader contextual factors.

To contextualize cybersecurity metrics, Bhol et al. [5] proposed a taxonomy categorizing metrics into vulnerabilities, protections, threats, users, and situations, leveraging Multi-Criteria Decision Making (MCDM) to enhance organizational cybersecurity. Similarly, Van Haastrecht et al. [45] reviewed socio-technical cybersecurity metrics for SMEs, introducing the SYMBALS method to prioritize and evaluate metrics, leading to a socio-technical framework tailored for SMEs. Further, Alim et al. [3] proposed metrics like *Individual Vulnerability* (VI) and *Relative Vulnerability* (VR) to assess social engineering risks in online networks. However, the frameworks in [5, 45] may not reflect the latest advancements and can lack responsiveness to evolving threats. Although the VI and VR [3] are innovative, they require validation and may oversimplify real-world network complexities, limiting their practical applicability.

These gaps highlight the need for an inclusive vulnerability metric considering victim characteristics, demographics, and attack-specific factors. Existing frameworks fail to capture the evolving nature of social cyber threats, underscoring the importance of specialized tools to better protect vulnerable populations.

### 2.2 AI-Driven Cybersecurity Solutions

Maddireddy and Maddireddy [29] proposed a multi-modal AI framework integrating machine learning, deep learning, and natural language processing (NLP) for ransomware detection. Evaluated on datasets like VirusShare and UNSW-NB15, it achieved high detection rates and adversarial robustness. Jamil et al. [24] introduced the MPMPA model to mitigate

phishing on Facebook, employing finite-state machines for detection and real-time alerts validated through realistic scenarios. Prasad et al. [35] developed *SnorCall*, using semi-supervised learning for analyzing illegal robocalls, achieving high labeling accuracy (90 - 100%) and insights into scam operations. These studies demonstrated innovative AI applications for specific cybersecurity threats. However, limitations exist. The framework in [29] faces a lack of scalability and interpretability, while MPMPA [24]’s focus on Facebook limits its generalizability. *SnorCall* [35]’s reliance on transient data and manual labeling introduced biases and ethical concerns. These limitations reflect the broader challenges in AI-driven cybersecurity, such as scalability, model interpretability, ethical considerations, and adaptability to dynamic threat landscapes, underscoring the need for further research and development to enhance robustness and usability.

### 2.3 Key Predictors of Vulnerability

Koning et al. [27] used Dutch survey data to examine socio-demographics, personality traits, and internet activities in fraud victimization, identifying higher risks for younger individuals, frequent internet users, and those with lower self-control, with openness to experience increasing susceptibility. Sur et al. [41] analyzed older adults, showing well-being and cognitive ability reduced fraud risk, while negative life events and loneliness increased it. Williams et al. [48] linked credit card fraud to financial confidence, training, marital status, and homeownership. Judges et al. [25] found lower cognitive ability and reduced honesty-humility as key predictors among older adults. However, their works are limited to reliance on post-fraud measurements [27], self-reported data [41, 48], and small, non-diverse samples [25]. These findings show the role of cognitive ability, personality traits, and socio-demographic factors in understanding vulnerability to social cyber threats, underscoring the need for broader, more robust studies.

### 2.4 Geographic, Demographic, and Temporal Dimensions of Social Scam Fraud

Edwards et al. [14] analyzed 5,402 online dating scam profiles, revealing Nigeria as a major source (30%) and identifying text/image reuse patterns via IP geolocation. Nguyen et al. [30] examined crime patterns in India using regression analysis, finding links between demographic factors and regional disparities. G. Nejad and Sabzian [18] investigated U.S. consumer financial fraud trends (2018–2022), noting spatial clustering of fraud, especially post-COVID-19, using techniques like Moran’s I. However, they could not resolve issues with dataset biases and proxy use [14], relied on secondary data lacking qualitative insights [30], and encountered underreporting and oversimplification in spatial analyses [18]. Despite these issues, they underscored the need for comprehensive methodologies to address fraud’s complex dynamics.

The Social Cyber Vulnerability Index (SCVI) introduced here bridges these gaps by integrating data sources like social media and surveys. It offers a context-aware framework to assess vulnerabilities across diverse contexts, transcending prior limitations in adaptability and inclusivity.

## 3 Measuring Social Cyber Vulnerabilities

### 3.1 Key Vulnerability Factors to Social Scams

Social scams represent a specialized category of online fraud [26], targeting victims through interactive channels such as email, websites, social media, and chat rooms. While online fraud typically leverages digital mediums to facilitate fraudulent activities, social scams focus on exploiting personal relationships, social dynamics, and trust [20]. This manipulation often involves deceptive communications or persuasive tactics, ultimately aiming to defraud individuals by undermining their behavioral and psychological defenses [42].

This work considers the following factors that shape an individual’s susceptibility to social cyberattacks. These factors align with the dimensions incorporated into the proposed *Social Cyber Vulnerability Index* (SCVI), which can be the basis for assessing risk to a given social scam:

1. **Individual Awareness and Knowledge:** This captures how well individuals understand social cyber threats and the protective measures available. Limited awareness and knowledge of specific scam types or associated security mechanisms heighten vulnerability [2].
2. **Behavioral Patterns:** This factor encompasses an individual’s online activities and security practices. Frequent exposure to risky platforms or inadequate protective behaviors can raise the likelihood of victimization [36].
3. **Psychological Factors:** This considers traits like trust and risk perception. Excessive trust in unfamiliar communications or a diminished sense of risk often translates into heightened susceptibility to scams [37].
4. **Past Experience:** This reflects prior encounters with social cyberattacks, including the individual’s responses and recovery strategies. Experiences can either strengthen resilience or, if inadequately addressed, leave persistent vulnerabilities [17].
5. **Frequency of Attacks:** This refers to the number of scam attempts a user encounters. A higher frequency increases the likelihood of at least one attack succeeding. Moreover, repeated scam exposure can desensitize individuals, potentially reducing their vigilance over time [47].
6. **Consequences of Attacks:** This involves significant financial or emotional harm that can exacerbate overall vulnerability, as high-stakes repercussions often leave lasting negative impacts [33].
7. **Sophistication of Attacks:** This assesses how convincingly a scam mimics legitimate communications or ex-

plotts personal information. Highly sophisticated (or realistic) social cyberattacks are more difficult to detect, elevating the victim’s susceptibility [11].

By examining these interrelated factors, the SCVI provides a holistic measure of an individual’s susceptibility to social cyberattacks. While numerous factors can contribute to vulnerabilities in social scams, this study focuses on seven key elements identified in existing research. We assess these factors using survey and social media datasets.

### 3.2 Social Cyber Vulnerability Index (SCVI)

We propose the *Social Cyber Vulnerability Index* (SCVI), assessing susceptibility to cyber social scams by evaluating individual vulnerability and attack severity factors.

The SCVI is formulated by:

$$SCVI_{i,k} = \alpha \cdot IVI_{i,k} + \beta \cdot ASI_{i,k}. \quad (1)$$

The  $IVI_{i,k}$  represents the Individual Vulnerability Index, indicating the vulnerability of an individual  $i$  to an attack  $k$ , while  $ASI_{i,k}$  denotes the Attack Severity Index, representing the impact of attack  $k$  on individual  $i$ . The weights  $\alpha$  and  $\beta$  are assigned to IVI and ASI, respectively, with  $\alpha + \beta = 1$ .

$IVI_{i,k}$  is given by:

$$IVI_{i,k} = w_{A_{i,k}} \cdot A_{i,k} + w_{B_{i,k}} \cdot B_{i,k} + w_{P_{i,k}} \cdot P_{i,k} + w_{E_{i,k}} \cdot E_{i,k}, \quad (2)$$

where  $A_{i,k} = A_{i,k}^A + A_{i,k}^K$ ,  $B_{i,k} = B_{i,k}^R + B_{i,k}^S$ ,  $P_{i,k} = P_{i,k}^C + P_{i,k}^I$ ,  $E_{i,k} = E_{i,k}^E + E_{i,k}^R$ , and  $w_{A_{i,k}} + w_{B_{i,k}} + w_{P_{i,k}} + w_{E_{i,k}} = 1$ .

The lack of individual  $i$ ’s **awareness and knowledge** ( $A_{i,k}$ ) includes unfamiliarity with social cyberattack  $k$  ( $A_{i,k}^A$ ) and knowledge of protective measures against  $k$  ( $A_{i,k}^K$ ). **Behavioral patterns** ( $B_{i,k}$ ) involve the frequency of risk-enhancing behaviors ( $B_{i,k}^R$ ) and the use of security practices ( $B_{i,k}^S$ ). **Psychological factors** ( $P_{i,k}$ ) include trust in communications related to  $k$  ( $P_{i,k}^C$ ) and risk perception and impulsivity ( $P_{i,k}^I$ ). **Experience** ( $E_{i,k}$ ) reflects past encounters with  $k$  ( $E_{i,k}^E$ ) and responses to such incidents ( $E_{i,k}^R$ ).

$ASI_{i,k}$  is defined by:

$$ASI_{i,k} = w_{F_{i,k}} \cdot F_{i,k} + w_{C_{i,k}} \cdot C_{i,k} + w_{S_{i,k}} \cdot S_{i,k}, \quad (3)$$

where  $F_{i,k} = F_{i,k}^{TA} + F_{i,k}^{AA}$ ,  $C_{i,k} = C_{i,k}^{FI} + C_{i,k}^{PI} + C_{i,k}^{SI}$ ,  $S_{i,k} = S_{i,k}^C + S_{i,k}^{SE}$ , and  $w_{F_{i,k}} + w_{C_{i,k}} + w_{S_{i,k}} = 1$ , respectively.

The impact of the **frequency of attack**  $k$  ( $F_{i,k}$ ) is determined by the frequency of attempted attacks ( $F_{i,k}^{TA}$ ) and the frequency of actual attacks encountered ( $F_{i,k}^{AA}$ ). The **consequence of a successful attack**  $k$  on individual  $i$  ( $C_{i,k}$ ) is represented by its financial impact ( $C_{i,k}^{FI}$ ), emotional or psychological impact ( $C_{i,k}^{PI}$ ), and impact on personal safety ( $C_{i,k}^{SI}$ ). The **sophistication of attack**  $k$  ( $S_{i,k}$ ) as perceived by individual  $i$

consists of the perceived degree to which the attack mimics legitimate communications ( $S_{i,k}^C$ ) and the use of personalized information or advanced social engineering techniques ( $S_{i,k}^{SE}$ ).

We consider the scale for each component (e.g.,  $A_{i,k}$ ) to be in  $[0, 5]$ , and its value will be represented as an integer. The  $IVI_{i,k}$  can be captured based on a Likert-scale (i.e., from 0 to 5) questionnaire or derived from an online user’s behavioral or lexical characteristics in social media. The  $ASI_{i,k}$  can be represented by the key characteristics of each attack perceived by online users [50]. The assigned weights for IVI and ASI balance individual vulnerability and attack severity, ensuring SCVI’s effectiveness for assessment and targeted interventions by integrating key factors across both dimensions.

## 4 Estimating SCVI

This section details the estimation of SCVI using two datasets: survey data from the iPoll dataset and social media data from Reddit scam reports. Integrating iPoll’s survey-based measures of risky behaviors and protective knowledge with Reddit’s user-generated scam reports, the SCVI achieves both statistical rigor and contextual depth, resulting in a comprehensive and adaptable measure of social cyber vulnerability. Together, these two datasets form a complementary foundation for SCVI estimation.

### 4.1 SCVI Using the iPoll Dataset

To compute the Social Cyber Vulnerability Index (SCVI), we utilize the survey data from the iPoll dataset [1]. The dataset, The Impostors: Stealing Money, Damaging Lives. An AARP National Survey of Adults 18+ (iPoll) thoroughly examines scam awareness, experiences, and behaviors among 4,596 adults across the United States. Conducted by NORC at the University of Chicago between January 2 and January 16, 2020. The dataset captures diverse aspects of online behavior, personal traits, and specific scams, such as romance scams, government impostor scams, and identity theft, across 113 variables. It employs computer-assisted telephone interviews (CATI) and web-based surveys, offering robust regional and national insights. While weighting factors are provided for analysis accuracy, the dataset highlights varying levels of scam susceptibility and awareness among different demographics, making it a critical resource for understanding and mitigating cyber vulnerabilities.

**Modeling the Individual Vulnerability Index (IVI).** IVI is modeled from the iPoll dataset by encoding and aggregating participant responses across seven dimensions reflecting cyber vulnerability. These dimensions include lack of Awareness ( $A_{i,k}^A$ ), lack of Knowledge of Protective Measures ( $A_{i,k}^K$ ), Frequency of Risk-Enhancing Behaviors ( $B_{i,k}^R$ ), Trust Level ( $P_{i,k}^C$ ), Risk Perception and Impulsivity ( $P_{i,k}^I$ ), Past Encounters



( $E_{i,k}^E$ ), and Responses to Past Incidents ( $E_{i,k}^R$ ). Each dimension is calculated by mapping survey responses to numerical scores based on predefined encoding schemes. For example, in  $A_{i,k}^A$ , a response of "Very concerned" is assigned a score of 0 (indicating high vulnerability), whereas "Not at all concerned" receives a score of 3. Similarly, responses like "False" in  $A_{i,k}^K$  are assigned a score of 5, reflecting higher vulnerability due to incorrect protective knowledge.

Each dimension is computed by averaging the encoded scores of relevant survey questions. The  $A_{i,k}^A$  factor measures familiarity-related concerns, while  $B_{i,k}^R$  captures the frequency of risk-enhancing behaviors, using scores derived from questions about daily or infrequent online activities. The Trust factor ( $P_{i,k}^C$ ) and Impulsivity factor ( $P_{i,k}^I$ ) are evaluated using responses about personal traits and decision-making tendencies, respectively. Encounters with cyber threats ( $E_{i,k}^E$ ) and responses to incidents ( $E_{i,k}^R$ ) are scored based on the frequency and severity of past experiences, including financial losses and emotional distress.

The final IVI score was computed as a weighted average of these factors by  $IVI = \frac{1}{7} (A_{i,k}^A + A_{i,k}^K + B_{i,k}^R + P_{i,k}^C + P_{i,k}^I + E_{i,k}^E + E_{i,k}^R)$ . This comprehensive index quantifies individual vulnerability factors based on familiarity, behavior, experience, and personality traits. This approach captures multidimensional aspects of vulnerability across participants.

**Modeling the Attack Severity Index (ASI).** ASI is derived from the iPoll dataset by encoding and aggregating survey responses across key dimensions: Frequency, Consequence, and Sophistication. Survey questions were mapped to numerical values using predefined encoding schemes, assigning higher scores to responses indicating greater risk or vulnerability.

Frequency ( $F_{i,k}$ ) was calculated as the sum of responses from questions measuring the prevalence of cyber events, with a score of 5 indicating frequent occurrences. Consequence ( $C_{i,k}$ ) aggregated scores from questions reflecting emotional, physical, or financial impacts, capturing outcome severity. Sophistication ( $S_{i,k}$ ) represented situational plausibility by summing scores from questions assessing the perceived legitimacy and sophistication of threats.

The ASI was computed as a weighted combination of these three components, with equal weights assigned to ensure a balanced vulnerability assessment. This comprehensive metric captures event frequency, outcome severity, and threat sophistication, offering insights into individual and systemic vulnerabilities. The encoded dataset facilitates further analysis across demographic and behavioral factors, enabling nuanced comparisons within the surveyed population.

The detailed feature extraction process for the IVI and ASI from the iPoll dataset is provided in Tables 4 and 5 in the appendices, respectively.

## 4.2 SCVI Using the Reddit Scam Reports

The Reddit Scam Reports dataset comprises user-generated posts from the subreddit r/scams, where individuals share experiences with scams. Spanning data from 2016 to 2023 (via the Pushshift API) and additional records from October to November 2024, the dataset focuses on scam reports for analysis. From an initial 5,000 observations, 450 scam reports were selected, ensuring equal representation across years.

Preprocessing steps included slang replacement, URL removal, and stripping mentions and hashtags. Text normalization involved converting to lowercase, expanding contractions, reducing elongated words, and removing non-alphanumeric characters while retaining punctuations. This cleaned dataset supports effective annotation of scam types and success metrics (successful or unsuccessful scams), providing a solid foundation for modeling and analyzing scam dynamics.

**Modeling the Individual Vulnerability Index (IVI).** The SCVI assesses individual vulnerabilities through key dimensions that capture user-specific factors influencing susceptibility to cyber threats. These dimensions offer a holistic view of personal risk, encompassing awareness, behaviors, psychological attributes, and prior experiences, collectively termed IVI. Below, we outline the core components of IVI:

- **Individual Lack of Awareness and Knowledge (A):** This measure evaluates users' exposure to scam-prevention strategies via fine-grained annotations of scam types and success. Participation in virtual communities improves well-being and knowledge sharing, reducing scam susceptibility [4]. Frequent engagement with scam discussions fosters familiarity with scams and avoidance methods, lowering vulnerability.
- **Behavioral Traits (B):** Behavioral vulnerability is modeled by analyzing user interaction frequency, such as posting behavior and linguistic markers. Frequent engagement with scam-related discussions, reflected in posts or reports, may indicate risk-taking or impulsive behavior linked to scam susceptibility [21]. Tools like Linguistic Inquiry and Word Count (LIWC) [34] assess markers such as clout scores, perceptual language, and informal markers [44]. For example, low clout scores suggest low confidence, while reliance on sensory descriptions indicates vulnerability to scams leveraging fake credibility cues. Behavioral traits like low self-confidence and over-reliance on perceptual processes significantly affect scam susceptibility [33]. Specific language patterns in online communities also influence susceptibility to scams and manipulative content [43], while skepticism, reflected in negations ("no," "not," "never"), helps resist misinformation and identify scams [49].
- **Psychological Factors (P):** Analytical thinking, emotional states, and personality traits significantly influence scam susceptibility. LIWC outputs, such as "affect" scores for emotional language and "cognitive processes" scores for

critical thinking, provide insights into these dimensions. Strong analytical and cognitive skills enhance scam resistance [23], while better cognitive abilities improve financial decision-making [19]. Emotional states like anxiety and neuroticism impair decision-making, increasing susceptibility [8, 10], with stress further exacerbating vulnerability [32]. Credulity, rather than general trust, is a specific fraud risk factor among older adults [39]. A psychological vulnerability score, derived using LIWC, identifies potential risks through markers like high emotional language, excessive exclamations, or overly positive tones ("posemo"), while skepticism and reflective thinking reduce vulnerability.

- **Experiences (E):** Prior exposure to scams is evaluated through annotations of posts detailing personal experiences, such as financial losses or emotional distress. This includes identifying whether users were scam victims. Such exposure may increase awareness or reinforce susceptibility [22, 40].

**Modeling the Attack Severity Index (ASI).** ASI is modeled by annotating scam reports for two critical attributes: *scam type* and *scam success*. The annotation process is conducted by two human annotators, with additional refinement using OpenAI’s API. Scam types include *phishing*, *investment scams*, *lottery scams*, *tech support scams*, *romance scams*, *online shopping scams*, *job scams*, and *undetected*. Scam success is represented as a binary variable (0 or 1). The ASI consists of three key components:

- **Frequency (F):** The frequency component quantifies the number of reported incidents for each scam type, providing insights into the prevalence of various scams [22].
- **Consequence (C):** The consequence of an attack assesses the impact of successful scams by analyzing financial losses [28] and emotional distress [38, 46]. Annotated reports capture monetary losses, while NLP techniques like emotion recognition models evaluate emotional impacts by aggregating positive and negative emotions in scam reports.
- **Sophistication (S):** The sophistication of an attack [13] measures the effectiveness of scams based on the proportion of successful scams relative to total reports within each type. Higher success rates indicate greater sophistication in deceiving victims, reflecting the severity of these scams.

## 5 Evaluation Setup

**Sensitivity Analysis of Weighting Schemes on SCVI Scores Using iPoll and Reddit Data.** A sensitivity analysis was conducted to evaluate the effect of varying weighting schemes on SCVI scores using two datasets: iPoll and Reddit scam reports. The analysis focused on the IVI components ( $w_A, w_B, w_P, w_E$ ) and ASI components ( $w_F, w_C, w_S$ ). The impact of each weight on the mean SCVI score and its variability was visualized and quantitatively assessed. This analysis provides insights into the relative importance of different

vulnerability factors and helps identify the most influential components driving SCVI variations.

**Monte Carlo Simulation for Analyzing Weight Variability and Uncertainty in SCVI Scores.** Monte Carlo simulations examined the uncertainty in SCVI scores by employing random sampling to model the effects of weight variability on the SCVI distribution. The plausible weight ranges for IVI ( $w_A, w_B, w_P, w_E$ ) and ASI ( $w_F, w_C, w_S$ ) were defined, ensuring that each index’s weights sum to one. Over 10,000 iterations, SCVI scores were recalculated, and aggregated results were analyzed to identify key patterns and weight configurations. This approach allows for a robust evaluation of potential fluctuations in SCVI scores, offering a deeper understanding of their stability under different weighting scenarios.

**Evaluation of SCVI Against CVSS and SVI.** The SCVI was evaluated against established indices, including the Common Vulnerability Scoring System (CVSS) [31] and the Social Vulnerability Index (SVI) [9]. CVSS is a widely used framework for rating software vulnerabilities, prioritizing management based on metrics such as Base, Temporal, and Environmental factors. Similarly, the SCVI employs metric-based mappings to assess fraud victimization, incorporating elements like attack vector, complexity, user interaction, and impacts on confidentiality, integrity, and availability.

The SVI quantifies community-level vulnerabilities to disasters and public health emergencies by focusing on sociodemographic and infrastructural factors. The SCVI extends these considerations by integrating survey data on individual behaviors, demographics, and cyber threat exposure, offering a composite view of individual vulnerability.

The evaluation involved calculating SCVI metrics and comparing them with CVSS and SVI scores derived from the iPoll dataset. This approach enabled a comprehensive analysis of SCVI’s capacity to integrate cyber-related behavioral data with sociodemographic factors, bridging established frameworks. Results are presented using tables to analyze significance, correlations, and outliers across demographic groups, including age, race-ethnicity, and gender.

## 6 Analyses of Evaluation Results

SCVI was estimated using the iPoll and Reddit datasets and compared with two existing metrics: SVI and CVSS.

### 6.1 Analyses of SCVI

**Analysis of the iPoll Dataset.** The analysis of the iPoll dataset reveals significant insights into user vulnerabilities and attack severity. Figure 1(a) illustrates the distribution of the IVI factors. The *Behavioral* factor exhibits a concentration of values around 1, indicating that most users engage minimally in behaviors that might increase their susceptibility

to scams. However, the *Psychological* factor shows a more even distribution, with most users scoring between 2 and 4, reflecting moderate psychological attributes influencing their vulnerability. The *Experience* factor is heavily skewed towards lower values, suggesting that many users have little to no prior exposure to scams. Meanwhile, the *Awareness and Knowledge* factor is widespread, with many users demonstrating low to moderate awareness of scams.

Figure 2(b) compares the IVI and ASI. The IVI distribution is between 2 and 3, reflecting moderate vulnerability levels among most users. In contrast, the ASI distribution is heavily skewed towards lower values, with a substantial frequency at 0, indicating that many users face attacks with minimal severity. However, a smaller yet significant group experiences higher ASI values around 4 and 5, highlighting the presence of severe attack cases. This disparity underscores the need to address user vulnerabilities and the impact of high-severity attacks to ensure comprehensive protection.

The factors contributing to the ASI provide additional insights, as shown in Figure 1(c). The *Frequency* factor reveals a bimodal distribution, with peaks near 0 and 5, indicating that users either rarely or frequently encounter scams. The *Consequence* factor is concentrated around moderate values, reflecting that the overall impact of scams is neither negligible nor catastrophic for most users. The *Sophistication* factor shows a diverse distribution, with peaks near both ends of the scale. This suggests a wide variation in the perceived authenticity and persuasiveness of scam attempts, with some being highly convincing while others are easily recognized.

In conclusion, the analysis of the iPoll dataset demonstrates a consistent pattern of low user vulnerability paired with high attack severity. The distribution of IVI and ASI factors suggests that while users may not frequently engage in risky behaviors, they remain vulnerable to severe and highly convincing scams. This highlights the need for targeted educational initiatives and awareness campaigns to reduce the impact of scams on vulnerable populations.

**Analysis of Reddit Dataset.** The distribution of the Individual Vulnerability Index (IVI) factors provides insights into user vulnerability, as shown in Figure 2(a). The *Behavioral* factor displays a strong skew towards lower values, indicating that most users exhibit low engagement in behaviors that might increase their susceptibility to scams. In contrast, the *Psychological* factor demonstrates a more balanced distribution, with most users scoring between 2 and 4, suggesting moderate psychological attributes contributing to vulnerability. The *Experience* factor is concentrated at discrete points, with a significant number of users reporting no prior scam encounters, while a smaller group shows high values, reflecting substantial past interactions with scams. Lastly, the *Awareness and Knowledge* factor is heavily skewed towards lower values, highlighting a general lack of awareness and knowledge about scams among users.

A comparison of the IVI and the Attack Severity Index (ASI) reveals a clear distinction in their distributions, as shown in Figure 2(b). The IVI primarily concentrates between 1 and 2, reflecting low to moderate vulnerability levels for most users. Conversely, the ASI distribution is shifted towards higher values, with the majority of users scoring between 3 and 4. This suggests that while users might exhibit relatively low vulnerability, the severity of the attacks they encounter is notably higher. This disparity underscores the need for targeted measures to bridge the gap between user vulnerability and the impact of the scams they experience.

The factors contributing to the ASI provide further insights into the nature of scams, as illustrated in Figure 2(c). The *Frequency* factor is highly polarized, with peaks near 0 and 5, suggesting that users experience either very frequent or infrequent attacks. The *Consequence* factor has a moderate peak around 3, indicating that the impact of scams is generally moderate for most users. However, the *Sophistication* factor, which measures the effectiveness of scams, displays a bimodal distribution with peaks near 2 and 4. This highlights varying levels of success and persuasiveness in scam attempts, with some being highly convincing while others are less so.

In conclusion, the analysis highlights a critical mismatch between user vulnerability and the severity of attacks. While most users exhibit low IVI scores, the high ASI values indicate that the attacks they encounter are often severe and impactful. Addressing this disparity requires enhancing user awareness and psychological resilience, particularly in recognizing and mitigating high-severity scams. These findings provide a foundation for designing effective strategies to reduce the impact of scams on vulnerable populations.

## 6.2 Sensitivity Analysis of IVI and ASI Components on SCVI Variability

The sensitivity analysis on the ipoll dataset reveals key trends in the contributions of IVI and ASI components to SCVI variability, as shown in Figure 4. In the IVI, the *Awareness* factor ( $w_A$ ) consistently increased SCVI scores, highlighting the exacerbating effect of lack of awareness. The *Experience* factor ( $w_E$ ) showed a decreasing trend with fluctuations, indicating interactions with other components. The *Psychological* factor ( $w_P$ ) strongly correlated with increased SCVI, while the *Behavioral* factor ( $w_B$ ) had minimal positive influence, reflecting limited impact in the Reddit context.

Regarding ASI components within the ipoll dataset, the *Frequency* factor ( $w_F$ ) showed a strong negative correlation with SCVI, indicating that lower frequencies of cyber-attacks increase vulnerability. The *Consequence* factor ( $w_C$ ) consistently decreased SCVI scores as its weight increased, reflecting the mitigating effect of preparedness against high-consequence events. The *Sophistication* factor ( $w_S$ ) displayed a mild positive trend with variability, suggesting its impact on SCVI depends on interactions with other factors.

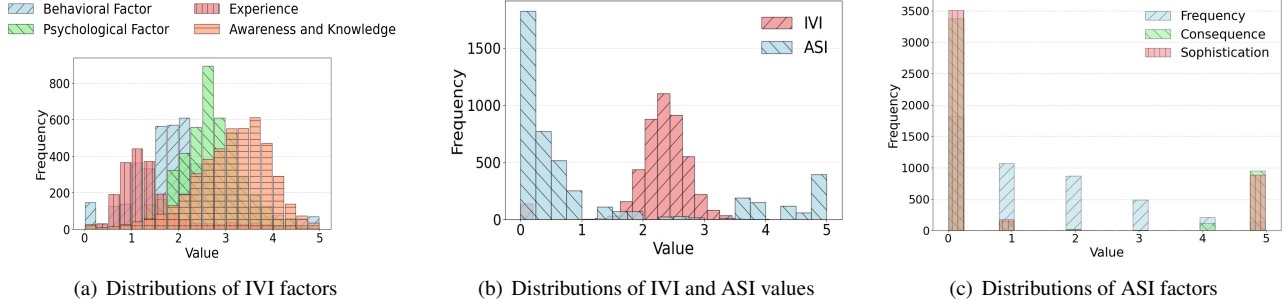


Figure 1: Comparison of IVI and ASI distributions and their contributing factors using the iPoll dataset.

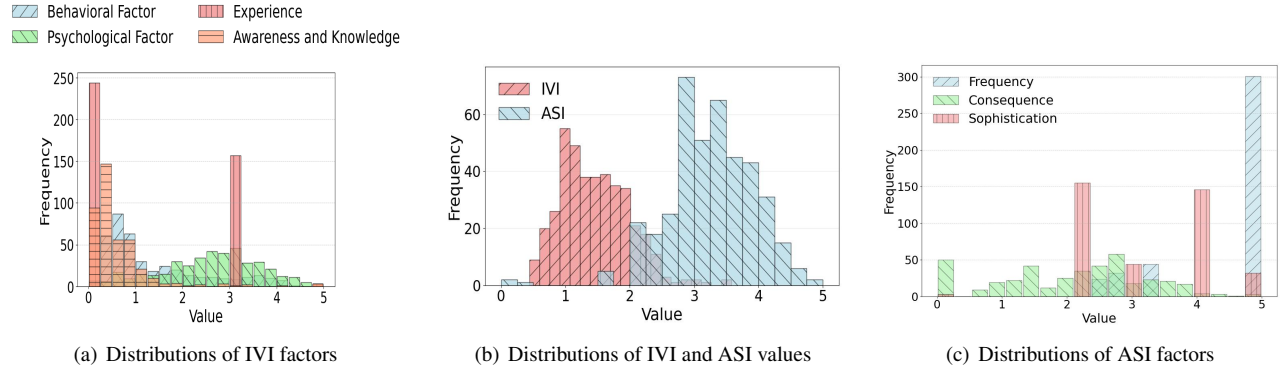


Figure 2: Comparison of IVI and ASI distributions and their contributing factors using the Reddit dataset.

The analysis of the Reddit dataset mirrored iPoll findings mostly but revealed differences in factor impacts, as in Figure 3. The strong exacerbating effects of higher weights in *Awareness* ( $w_A$ ) and *mitigating effects of Experience* ( $w_E$ ) on SCVI were consistent across both datasets. In contrast, the *Psychological* factor ( $w_P$ ) exhibited a slight negative driver of SCVI. The *Behavioral* factor ( $w_B$ ) showed a slight negative correlation with SCVI in the Reddit dataset, in comparison with the iPoll analysis. For the ASI components, *Frequency* ( $w_F$ ) was again a major contributor to SCVI in both datasets, though its impact was stronger in the Reddit data. The *Consequence* factor ( $w_C$ ) similarly reduced SCVI across both datasets but had a more pronounced effect in the Reddit dataset. The *Sophistication* factor ( $w_S$ ) had a more consistent and significant positive impact on SCVI in the iPoll dataset compared to its less stable influence in the Reddit dataset.

These analyses underscore the robust yet context-sensitive nature of SCVI components across different datasets. The universal impacts of Awareness, Experience, Consequence and Frequency on SCVI were evident. Nonetheless, variations in the magnitude and direction of relationships for Psychological, Behavioral and Sophistication factors across datasets highlight the influence of data-specific attributes. Further investigations into additional datasets and varying contexts can enhance the applicability and robustness of SCVI as a comprehensive

metric for assessing cyber vulnerabilities. We leave this for our future research.

### 6.3 Effect of Weight Variability in SCVI

The SCVI values and corresponding weight configurations were aggregated across multiple iterations to identify key patterns, as shown in Figures 5(a) and 5(b). The results highlighted two primary peaks shown in Figure 5(a) for the iPoll dataset:

**Primary Peak (Group 1):** SCVI was predominantly driven by Experience ( $w_E$ ) in the Individual Vulnerability Index (IVI) and Sophistication ( $w_S$ ) in the Attack Severity Index (ASI). Experience made the largest contribution to IVI, with a mean value of 0.421 and a standard deviation of 0.084. Similarly, Sophistication dominated ASI contributions, with a mean value of 0.447 and a standard deviation of 0.095. These findings emphasize the importance of systemic Sophistication and individual experience in influencing SCVI scores.

**Secondary Peak (Group 2):** SCVI values were primarily influenced by Awareness ( $w_A$ ) and Psychological Factors ( $w_P$ ) in IVI, as well as Frequency ( $w_F$ ) in ASI. Awareness emerged as a significant IVI contributor (Mean: 0.372, Std: 0.026), while Frequency was the dominant ASI factor (Mean: 0.436, Std: 0.033). In this group, Sophistication and Experi-



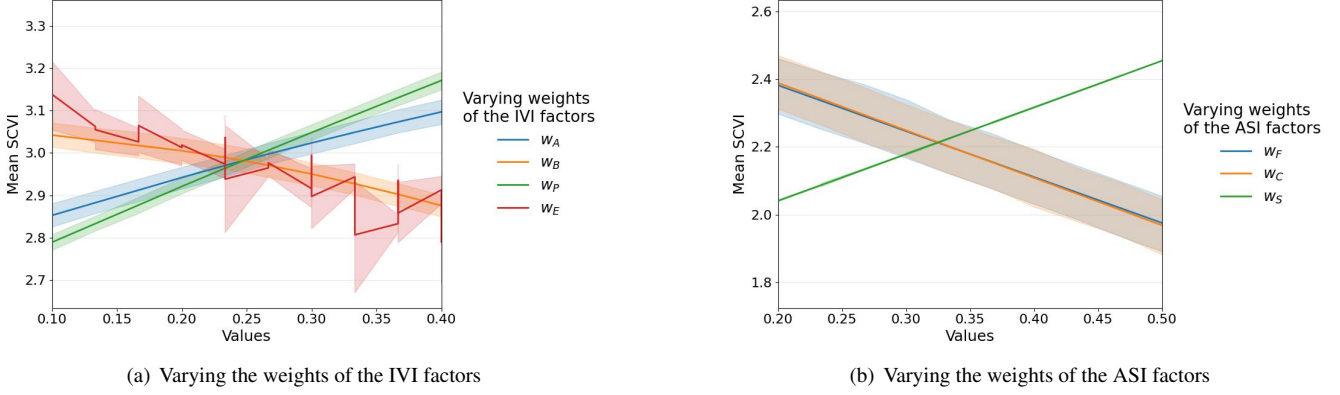


Figure 3: Sensitivity analysis of the iPoll dataset for IVI and ASI factors. Note that  $w_A$ ,  $w_B$ ,  $w_P$ , and  $w_E$  are the weights for ‘Awareness,’ ‘Behavioral,’ ‘Psychological,’ and ‘Experience’ factors in the individual vulnerability index (IVI) correspondingly.  $w_F$ ,  $w_C$ , and  $w_S$  refer to ‘Frequency,’ ‘Consequence,’ and ‘Sophistication’ in the Attack Security Index (ASI), respectively.

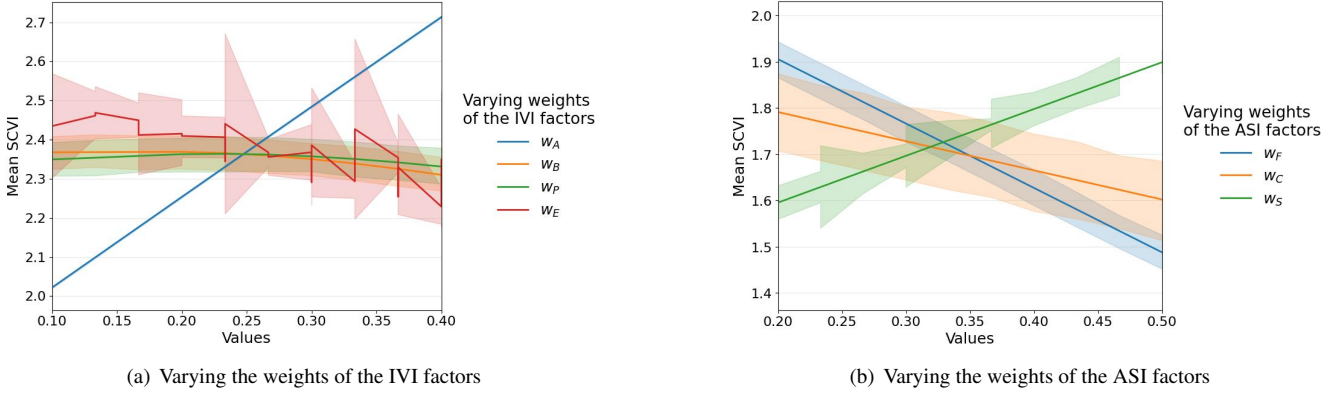


Figure 4: Sensitivity analysis of the Reddit dataset for IVI and ASI factors. Note that  $w_A$ ,  $w_B$ ,  $w_P$ , and  $w_E$  are the weights for ‘Awareness,’ ‘Behavioral,’ ‘Psychological,’ and ‘Experience’ factors in the individual vulnerability index (IVI) correspondingly.  $w_F$ ,  $w_C$ , and  $w_S$  refer to ‘Frequency,’ ‘Consequence,’ and ‘Sophistication’ in the Attack Security Index (ASI), respectively.

ence played a minimal role, with SCVI relying more on attack frequency, consequences, and psychological factors.

We find the following observation in Figure 5(b), which analyzed the Reddit dataset.

**Low SCVI Outliers:** SCVI values exhibited low variability, ranging from 1.039 to 1.098, indicating consistency. Behavioral Factors ( $w_B$ ) emerged as the dominant contributor, ranging from 0.33 to 0.37, with Experience ( $w_E$ ) also playing a significant role in certain cases.  $\alpha$  values ranged between 0.562 and 0.593, slightly favoring IVI, while  $\beta$  values from 0.406 to 0.437 represented a balanced ASI contribution. This indicates that user behaviors, such as risk-taking tendencies and prior experiences, significantly influence SCVI, while the overall stability of the metric reflects a robust model for assessing vulnerabilities.

**High SCVI Outliers:** SCVI values showed greater variability, ranging from 2.182 to 2.211. Experience ( $w_E$ ) and

Frequency ( $w_F$ ) were the most influential factors, consistently exceeding 0.45, highlighting the importance of users’ prior exposure to scams and the prevalence of cyberattack attempts in shaping vulnerabilities. and  $\alpha$  values between 0.403 and 0.419 suggested a balanced primary contribution from IVI components, while  $\beta$  values between 0.580 and 0.596 indicated a stronger impact from ASI components. This distribution suggests that while individual vulnerability factors contribute significantly, the attack-specific characteristics, particularly frequency, dominate overall SCVI scores.

**Comparative Analysis:** Both datasets highlight the dominant roles of Experience ( $w_E$ ) and Frequency ( $w_F$ ) in driving SCVI scores. However, the Reddit dataset emphasized Behavioral Factors ( $w_B$ ) in low SCVI cases and a more pronounced influence of Frequency in high SCVI cases, suggesting that user behaviors are more critical in reducing vulnerability, while the prevalence of cyberattacks becomes a stronger driver

of high SCVI scores. In contrast, the iPoll dataset demonstrated stronger contributions from Sophistication ( $W_S$ ) and Awareness ( $w_A$ ), reflecting the importance of user Awareness and the complexity of attacks in shaping vulnerabilities. This comparison underscores the contextual differences between datasets, highlighting the varying impacts of individual behaviors and attack characteristics on SCVI.

Summarizing the insights above, while SCVI components exhibit universal trends, their relative contributions differ significantly across datasets, reflecting the contextual characteristics of the data. For instance, the iPoll dataset highlights the dominant roles of Sophistication ( $W_S$ ) and Awareness ( $w_A$ ), emphasizing the importance of attack complexity and user awareness in shaping vulnerabilities. Conversely, the Reddit dataset underscores the critical influence of Behavioral Factors ( $w_B$ ) in low SCVI cases and Frequency ( $w_F$ ) in high SCVI cases, illustrating the varying impacts of user behaviors and cyberattack prevalence.

This variability highlights that tailoring SCVI metrics is critical to the specific attributes of the analyzed dataset. Future research could delve deeper into these dynamics to enhance SCVI’s adaptability and robustness, enabling more accurate assessments of vulnerabilities across diverse environments.

#### 6.4 Comparative Analysis with SVI and CVSS Across Demographic Groups

The SCVI metric effectively captures individual-level vulnerabilities in social cyber contexts, surpassing SVI and CVSS through its integration of socio-demographic, behavioral, and cyber-specific dimensions. As shown in Figure 6, correlation analysis reveals a moderate positive correlation with CVSS (Spearman = 0.33,  $p = 0.0$ ), indicating alignment in identifying technological vulnerabilities. In contrast, SCVI exhibits a weaker correlation with SVI ( $-0.01$ ,  $p = 0.4836$ ), reflecting SVI’s broader socio-environmental focus, which lacks specificity in cyber-related risks.

SCVI complements CVSS and SVI by incorporating dimensions that address individual vulnerabilities in socio-cyber contexts. Its alignment with CVSS and divergence from SVI highlight its ability to integrate social and behavioral factors, making it a vital tool for inclusive cybersecurity strategies.

Analysis of trends across gender, race-ethnicity, and age groups reveals consistent patterns in vulnerability metrics. Table 1 shows that the male group has the highest SVI, CVSS, and SCVI values at 2.57, 3.52, and 3.47, respectively. The SVI difference between male and female groups is relatively minimal (2.36%), while CVSS and SCVI show significantly larger gaps at 9.84% and 16.88%, potentially due to exposure to riskier technological environments.

As shown in Table 2, among racial and ethnic groups, the White, non-Hispanic group has the lowest values across all indices, with SVI at 2.25 and CVSS at 3.02. Conversely, the Hispanic group records the highest values, with SVI, CVSS,

Table 1: DISTRIBUTION OF SVI, CVSS, AND SCVI METRICS BY GENDER

Gender	SVI	CVSS	SCVI
Female	2.51	3.19	2.93
Male	2.57	3.52	3.47

Table 2: DISTRIBUTION OF SVI, CVSS, AND SCVI METRICS BY RACE-ETHNICITY

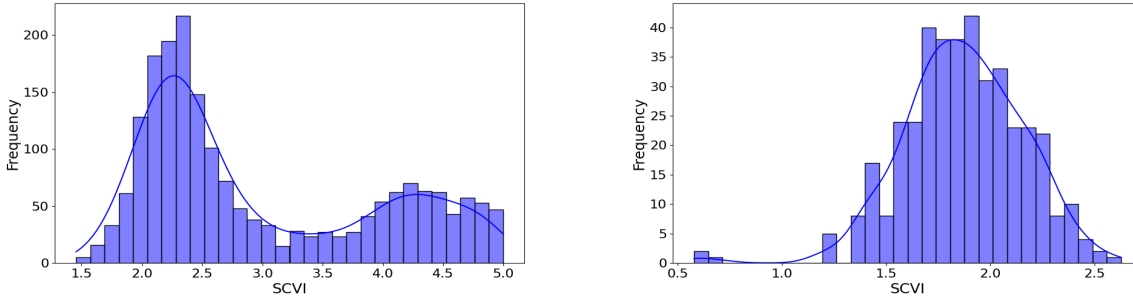
Race-Ethnicity	SVI	CVSS	SCVI
White, non-Hispanic	2.25	3.31	3.02
Black, non-Hispanic	3.34	3.34	3.40
Other, non-Hispanic	3.01	3.74	3.33
2+, non-Hispanic	2.89	3.32	3.14
Asian, non-Hispanic	2.90	3.32	3.02
Hispanic	3.72	4.06	3.79

and SCVI at 3.72, 3.79, and 4.06, respectively. SCVI generally exceeds CVSS, which in turn exceeds SVI. These findings suggest that targeted cybersecurity programs for Hispanic groups could address their higher vulnerability. In contrast, the lower scores for White, non-Hispanic groups may reflect better access to resources, education, or programs.

Holistically, the three columns in Table 2 showcase a trend of SVI scores almost always being lower than that of CVSS or SCVI. In fact, the only two groups where SVI is not the lowest is: black, non-Hispanic (by 0.3%) and other, non-Hispanic (by 10%) race-ethnicity groups in the table. No group displayed SVI having the highest vulnerability index.

In age group analysis, presented in Table 3, younger demographics (18-24 and 25-29) are overall more vulnerable than older age groups with SCVI being the highest. This indicates a stronger susceptibility to social cyber vulnerabilities possibly due to higher technology usage. As age increases, all metrics notably decrease with middle-aged and older-aged groups’ CVSS scores leading the other two metrics. Note that there is a slight increase in all metrics for the 45-49 and 50-54 age groups, perhaps indicating higher engagement but less awareness with technology and scams than neighboring age groups. SVI scores are also the lowest across all age groups.

These findings validate SCVI as a robust and comprehensive metric, surpassing traditional indices in capturing nuanced social cyber vulnerabilities. SCVI consistently exhibits the highest scores across demographic categories, emphasizing its utility in identifying risks associated with technological behaviors and social factors. This distinction highlights its potential as a comprehensive tool for addressing individual-level vulnerabilities, particularly in demographic groups with higher susceptibility to cyber threats.



(a) Monte Carlo Analysis of SCVI Components in the iPoll Dataset: Primary and Secondary Peaks. (b) Monte Carlo Analysis of SCVI Components in the Reddit Dataset: Low and High SCVI Outliers.

Figure 5: Monte Carlo analysis of SCVI components in the iPoll and Reddit datasets.

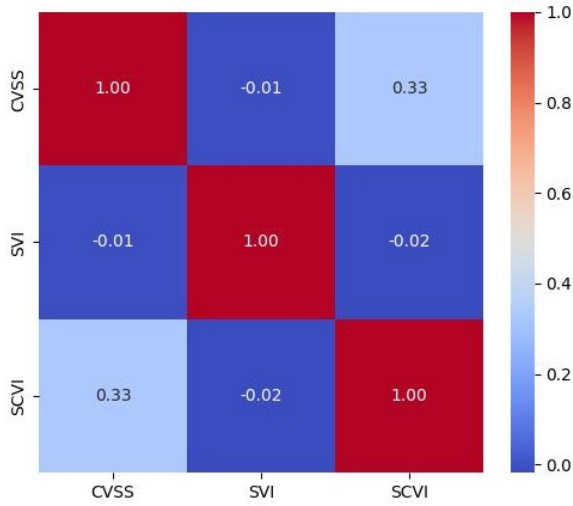


Figure 6: Spearman correlation heatmap across SCVI, SVI, and CVSS metrics.

## 6.5 Regional Disparities in Cyber Vulnerability: Analysis of SCVI Across U.S. States Using the iPoll Dataset

The SCVI calculated from the iPoll dataset analyses regional disparities in cyber vulnerability across the United States. The dataset includes mean SCVI scores, confidence intervals, and sample sizes for each state, providing critical insights into the geographical distribution of cyber threats and the effectiveness of local cybersecurity measures.

Data was visualized through a heatmap to illustrate the SCVI distribution, with color intensities adjusted based on sample sizes to reflect data reliability in Figure 7. Table 6 in the appendices provides a detailed summary of the state-wise SCVI scores, sample sizes, and confidence intervals.

States like Alaska, Rhode Island, and Nevada, which exhibit higher mean SCVI scores and smaller sample sizes, suggest an

Table 3: DISTRIBUTION OF SVI, CVSS, AND SCVI METRICS BY AGE GROUPS

Age Group	SVI	CVSS	SCVI
18-24	3.85	4.69	4.14
25-29	3.11	3.59	3.50
30-44	3.31	3.33	3.11
45-49	2.74	3.82	3.45
50-54	2.42	3.81	3.45
55-64	2.25	3.24	2.95
65+	2.12	2.82	2.63

elevated risk level potentially influenced by regional factors. The wide confidence intervals in these measurements indicate a significant uncertainty in the SCVI estimates, attributed to the limited data availability.

In contrast, populous states such as California, Texas, and New York demonstrate lower and more stable SCVI scores, suggesting either lower cyber vulnerability or more effective cybersecurity practices. The robustness of data from these states provides clearer insights into their cyber health landscapes. The variability in SCVI scores between states like Connecticut and Pennsylvania, despite similar sample sizes, underscores the impact of socio-economic or infrastructural factors on cyber vulnerability. This variability highlights the complex dynamics affecting cybersecurity across regions. These observations align with the Federal Trade Commission’s (FTC) fraud reports in the 2020 FTC Data Book [12], which highlight similar regional trends in cyber fraud incidents and vulnerabilities.

The inverse relationship observed between the mean SCVI scores and sample sizes across several states suggests that smaller samples may capture specific regional extremes or anomalies not present in larger, more representative data sets. This observation necessitates cautious interpretation of data and, possibly, further investigation into these regions.

The analysis underscores the need for region-specific cyber





dynamics influencing cyber vulnerabilities.

**Fifth**, the cross-sectional nature of the datasets further limits the ability to draw causal interpretations and obscures how vulnerabilities evolve. Longitudinal studies are needed to capture dynamic changes in cyber vulnerabilities and provide deeper insights into temporal trends. Furthermore, sociocultural variations in trust cues and persuasion tactics restrict the generalizability of SCVI to diverse populations. Addressing these **cultural specificities** in future research is essential to enhance the framework’s applicability across global contexts.

**Lastly**, the **continuous evolution of cybercriminal tactics** necessitates regular updates to the SCVI framework to incorporate emerging scam vectors and adapt to the rapidly changing cyber landscape.

These limitations highlight critical areas for refinement and underscore the need for future research to improve the SCVI framework’s adaptability, scalability, and robustness as a comprehensive metric for assessing social cyber vulnerabilities.

### 7.3 Future Work

To address the identified limitations of the **SCVI framework** and build on its strengths, several key future research directions are proposed.

**First**, the SCVI offers a holistic perspective on cyber vulnerability by integrating individual-level risk factors with contextual attack severity. Results from the iPoll and Reddit datasets illustrate that while many individuals maintain moderate or even low self-assessed susceptibility, they often encounter sophisticated or severe scams. This mismatch underscores **the need for targeted interventions** addressing user vulnerability and the high-impact nature of many prevalent cyber threats. SCVI’s modular design, encompassing dimensions such as awareness, psychological traits, past experiences, and scam sophistication, facilitates more robust analyses than traditional metrics like CVSS and SVI, primarily focusing on software exploits or broad social indicators.

**Second**, efforts should improve **the reliability of input data** by reducing biases associated with self-reported data. Incorporating alternative data collection methods, such as behavioral tracking or third-party validation of reported incidents, can help mitigate recall bias and under-reporting. These methods would provide a more objective foundation for assessing vulnerabilities and enhancing SCVI’s robustness.

**Third**, **expanding the geographic and demographic scope of SCVI** is essential for capturing a more diverse range of cultures, languages, and age groups. This would improve the framework’s applicability and inclusivity. Leveraging diverse datasets across cultural and linguistic contexts could enhance SCVI’s generalizability. Moreover, addressing underrepresented populations, such as older adults with limited Internet access and non-English speakers, would ensure a more comprehensive analysis of cyber vulnerabilities.

**Fourth**, **addressing sample size variability**, particularly in geographical analyses, is crucial for improving the reliability of SCVI estimates. Increasing sample sizes for underrepresented regions and states would reduce confidence intervals and enhance regional assessments. Stratified sampling methods or region-specific weighting schemes could also refine regional vulnerability analyses.

**Fifth**, **longitudinal studies** are needed to track individual and community-level changes in vulnerability over time. Such studies would provide valuable insights into how behaviors, awareness levels, and scam tactics evolve in response to emerging threats and mitigation strategies. Temporal analyses could facilitate the identification of trends and inform proactive intervention strategies.

**Sixth**, **developing data-driven weighting methods** represents a promising direction for optimizing SCVI. Incorporating machine learning techniques, such as feature importance estimation or optimization algorithms, would enable dynamic adjustments of IVI and ASI weights based on real-time threat intelligence. These approaches would enhance SCVI’s predictive power and accuracy.

**Seventh**, **regular updates to SCVI** are critical to maintaining its relevance in an ever-evolving digital threat landscape. This includes incorporating emerging scam typologies such as AI-powered phishing and deepfake-related fraud. By adapting to new threats, SCVI can remain a versatile and contextually relevant framework for assessing cyber vulnerabilities.

**Lastly**, **addressing cultural specificity** is essential for enhancing SCVI’s global applicability. Sociocultural differences in trust cues, communication styles, and persuasion tactics must be explored to localize the framework. Comparative studies across regions provide deeper insights into cultural and behavioral factors, further refining SCVI.

These future research directions aim to enhance the SCVI framework’s adaptability, robustness, and inclusivity. By expanding its geographic and demographic scope, integrating data-driven methods, and continuously updating the framework to reflect emerging threats, SCVI can become an indispensable tool for assessing and mitigating social cyber vulnerabilities in diverse and dynamic contexts.

## Ethics Considerations

This research was conducted with careful attention to ethical principles, particularly regarding data privacy and the responsible use of publicly available datasets.

1. **Use of the iPoll Dataset [1]:** The iPoll dataset, which was utilized in this study, is publicly available and has already been anonymized to ensure that no personally identifiable information (PII) is present. This guarantees compliance with data privacy standards and ethical guidelines for the use of secondary data.
2. **Reddit Scam Reports:** The Reddit scam reports used in this study were collected via web crawling from publicly accessible forums where individual users share experiences with scams. The user identifiers in the crawled data were anonymized, and no attempts were made to de-anonymize or infer personal information about users. This aligns with ethical research practices and Reddit's terms of service regarding data usage.
3. **Compliance with Data Policies:** We ensured that our data collection and analysis process complied with the terms of use for both datasets. Additionally, care was taken to avoid data misuse that could harm individuals or groups.
4. **Research Transparency and Reproducibility:** To promote transparency and reproducibility, we documented our data collection and preprocessing methods in detail while ensuring that any data shared as part of this research does not compromise the privacy or anonymity of individuals.

By adhering to these measures, this study upholds the ethical standards expected in academic research and respects the privacy and rights of individuals whose information is included in the datasets.

## Compliance with the Open Science Policy

This study adheres to the Open Science Policy by ensuring transparency, accessibility, and reproducibility in all aspects of the research process. Below are the key measures undertaken:

**Data Transparency and Availability:** The datasets used in this study include the publicly accessible iPoll dataset and Reddit scam reports. The iPoll dataset is available from the Roper Center for Public Opinion Research and has been anonymized to remove personally identifiable information. The Reddit data, collected through web crawling, consists of user-generated content shared in public forums and was processed to ensure no privacy violations. Detailed descriptions of these datasets, including preprocessing steps, are provided in the appendices.

**Methodological Reproducibility:** To promote reproducibility, all analytical methods, modeling processes, and evaluation techniques employed in this study are detailed.

The computational pipeline used to estimate the Social Cyber Vulnerability Index (SCVI) has been documented comprehensively, including sensitivity analysis and Monte Carlo simulations.

**Code Accessibility:** The codebase developed for calculating the SCVI, performing sensitivity analysis, and generating the results presented in this paper will be made available upon acceptance of this paper. It will be hosted on a public repository (e.g., GitHub) with proper documentation to facilitate reuse and further research.

**Ethical Compliance:** Ethical considerations, including data anonymization and adherence to data usage policies, were strictly followed. The details of these measures are outlined in the Ethics Considerations section.

By complying with these principles, this work aligns with the Open Science Policy, ensuring that the research is accessible, transparent, and reproducible, thereby fostering broader collaboration and advancing the field of cybersecurity research.

## References

- [1] AARP, "The Impostors: Stealing Money, Damaging Lives. An AARP National Survey of Adults 18+, 2020 [Dataset]," 2020, roper #31119515, Version 2. NORC at the University of Chicago [producer]. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research [distributor]. Access Date: Aug-30-2024.
- [2] S. M. Albladi and G. R. Weir, "Predicting individuals' vulnerability to social engineering in social networks," *Cybersecurity*, vol. 3, no. 1, p. 7, 2020.
- [3] S. Alim, D. Neagu, and M. Ridley, "Axioms for vulnerability measurement of online social network profiles," in *International Conference on Information Society (i-Society 2011)*. IEEE, 2011, pp. 241–247.
- [4] Z. Ayachi and R. Jallouli, "Virtual communities and wellbeing: A systematic literature review and recommendations for future research," in *International Conference on Digital Economy*. Springer, 2021, pp. 64–86.
- [5] S. G. Bhol, J. Mohanty, and P. K. Pattnaik, "Taxonomy of cyber security metrics to measure strength of cyber security," *Materials Today: Proceedings*, vol. 80, pp. 2274–2279, 2023.
- [6] P. E. Black, K. Scarfone, M. Souppaya *et al.*, "Cyber security metrics and measures," *Wiley Handbook of Science and Technology for Homeland Security*, pp. 1–15, 2008.
- [7] M. Bolpagni, "Cyber risk index: a socio-technical composite index for assessing risk of cyber attacks with

- negative outcome,” *Quality & Quantity*, vol. 56, no. 3, pp. 1643–1659, 2022.
- [8] P. A. Boyle, L. Yu, G. Mottola, K. Innes, and D. A. Bennett, “Degraded rationality and suboptimal decision-making in old age: A silent epidemic with major economic and public health implications,” *Public Policy & Aging Report*, vol. 32, no. 2, pp. 45–50, 2022.
- [9] CDC/ATSDR, *Social Vulnerability Index*, 2025, accessed: January 17, 2025. [Online]. Available: <https://www.atsdr.cdc.gov/place-health/php/svi/index.html>
- [10] J.-H. Cho, H. Cam, and A. Oltramari, “Effect of personality traits on trust and risk to phishing vulnerability: Modeling and analysis,” in *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*. IEEE, 2016, pp. 7–13.
- [11] B. Collier and R. Clayton, “A “sophisticated attack”? innovation technical sophistication and creativity in the cybercrime ecosystem,” in *The 21st Workshop on the Economics of Information*, 2022.
- [12] F. T. Commission, “Consumer sentinel network data book 2020,” 2021, accessed: 2025-01-21. [Online]. Available: [https://www.ftc.gov/system/files/documents/reports/consumer-sentinel-network-data-book-2020/csn\\_annual\\_data\\_book\\_2020.pdf](https://www.ftc.gov/system/files/documents/reports/consumer-sentinel-network-data-book-2020/csn_annual_data_book_2020.pdf)
- [13] A. A. Darem, T. M. Alkhaldi, M. Alahmari, A. A. Alhashmi, A. M. Alashjaee, S. M. Alanazi, and S. A. Ebad, “Beyond technical barriers: A multidimensional conceptual framework for understanding and countering cyber scam susceptibility,” *International Journal of Human-Computer Interaction*, pp. 1–26, 2024.
- [14] M. Edwards, G. Suaraz-Tangil, C. Peersman, G. Stringhini, A. Rashid, and M. Whitty, “The geography of online dating fraud,” in *Workshop on Technology and Consumer Protection 2018*. IEEE, Institute of Electrical and Electronics Engineers, 2018.
- [15] B. E. Flanagan, E. W. Gregory, E. J. Hallisey, J. L. Heitgerd, and B. Lewis, “A social vulnerability index for disaster management,” *Journal of Homeland Security and Emergency Management*, vol. 8, no. 1, p. 0000102202154773551792, 2011.
- [16] E. D. Frauenstein and S. Flowerday, “Susceptibility to phishing on social network sites: A personality information processing model,” *Computers & Security*, vol. 94, p. 101862, 2020.
- [17] S. S. Fredrick, L. N. Jenkins, and C. M. Dexter, “Resiliency in young adulthood and associations among retrospective peer victimization and internalizing problems,” *Journal of Child & Adolescent Trauma*, vol. 14, no. 3, pp. 367–379, 2021.
- [18] M. G. Nejad and H. Sabzian, “Spatiotemporal patterns of consumer financial fraud in the united states,” *International Journal of Bank Marketing*, 2024.
- [19] K. J. Gamble, P. A. Boyle, L. Yu, and D. A. Bennett, “Aging and financial decision making,” *Management science*, vol. 61, no. 11, pp. 2603–2610, 2015.
- [20] Z. Guo, J.-H. Cho, R. Chen, S. Sengupta, M. Hong, and T. Mitra, “Online social deception and its countermeasures: A survey,” *IEEE Access*, vol. 9, pp. 1770–1806, 2020.
- [21] A. M. Herman, H. D. Critchley, and T. Duka, “Risk-taking and impulsivity: The role of mood states and interoception,” *Frontiers in Psychology*, vol. 9, p. 1625, 2018.
- [22] M. Houtti, A. Roy, V. N. R. Gangula, and A. Walker, “A survey of scam exposure, victimization, types, vectors, and reporting in 12 countries,” *Journal of Online Trust and Safety*, vol. 2, no. 4, 2024.
- [23] T. M. Hruschka and M. Appel, “Learning about informal fallacies and the detection of fake news: An experimental intervention,” *PLoS One*, vol. 18, no. 3, p. e0283238, 2023.
- [24] A. Jamil, K. Asif, Z. Ghulam, M. K. Nazir, S. M. Alam, and R. Ashraf, “MPMPA: A mitigation and prevention model for social engineering-based phishing attacks on Facebook,” in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 5040–5048.
- [25] R. A. Judges, S. N. Gallant, L. Yang, and K. Lee, “The role of cognition, personality, and trust in fraud victimization in older adults,” *Frontiers in Psychology*, vol. 8, p. 588, 2017.
- [26] J. Kävrestad and M. Nohlberg, “Defining and modelling the online fraud process,” in *HAISA*, 2018, pp. 203–213.
- [27] L. Koning, M. Junger, and B. Veldkamp, “Risk factors for fraud victimization: The role of socio-demographics, personality, mental, general, and cognitive health, activities, and fraud knowledge,” *International Review of Victimology*, p. 02697580231215839, 2023.
- [28] Z. Lwin Tun and D. Birks, “Supporting crime script analyses of scams with natural language processing,” *Crime Science*, vol. 12, no. 1, p. 1, 2023.
- [29] B. R. Maddireddy and B. R. Maddireddy, “Evolutionary algorithms in ai-driven cybersecurity solutions for

- adaptive threat mitigation,” *International Journal of Advanced Engineering Technologies and Innovations*, vol. 1, no. 2, pp. 17–43, 2021.
- [30] N. T. Nguyen, M. E. Lourens, R. Manjre, V. Prakash, S. Patil, and M. R. Kamaluddin, “Analysis of social crime patterns in regions based on demographic (geographical) distribution,” *Review of International Geographical Education Online*, vol. 11, no. 7, 2021.
- [31] NIST, *NVD - Vulnerability Metrics: Common Vulnerability Scoring System (CVSS)*, 2025, accessed: January 17, 2025. [Online]. Available: <https://nvd.nist.gov/vuln-metrics/cvss>
- [32] G. Norris and A. Brookes, “Personality, emotion and individual differences in response to online fraud,” *Personality and Individual Differences*, vol. 169, p. 109847, 2021.
- [33] G. Norris, A. Brookes, and D. Dowell, “The psychology of internet fraud victimisation: A systematic review,” *Journal of Police and Criminal Psychology*, vol. 34, pp. 231–245, 2019.
- [34] I. Pennebaker Conglomerates, *Linguistic Inquiry and Word Count (LIWC) [Computer software]*, 2022, available from <https://www.liwc.app/>.
- [35] S. Prasad, T. Dunlap, A. Ross, and B. Reaves, “Diving into robocall content with snorcall,” in *The 32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 427–444.
- [36] B. W. Reyns and B. Henson, “Security in a digital world: Understanding and preventing cybercrime victimization,” *Security Journal*, vol. 26, pp. 311–314, 2013.
- [37] C. A. Robb and S. Wendel, “Who can you trust? assessing vulnerability to digital imposter scams,” *Journal of Consumer Policy*, vol. 46, no. 1, pp. 27–51, 2023. [Online]. Available: <https://link.springer.com/article/10.1007/s10603-022-09531-6>
- [38] F. Safari and A. Chalechale, “Emotion and personality analysis and detection using natural language processing, advances, challenges and future scope,” *Artificial Intelligence Review*, vol. 56, no. Suppl 3, pp. 3273–3297, 2023.
- [39] J. Shao, W. Du, T. Lin, X. Li, J. Li, and H. Lei, “Credulity rather than general trust may increase vulnerability to fraud in older adults: A moderated mediation model,” *Journal of Elder Abuse & Neglect*, vol. 31, no. 2, pp. 146–162, 2019.
- [40] S. Sheng, M. Holbrook, P. Kumaraguru, L. F. Cranor, and J. Downs, “Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 373–382.
- [41] A. Sur, M. DeLiema, and E. Brown, “Contextual and social predictors of scam susceptibility and fraud victimization,” *Available at SSRN 4053903*, 2021.
- [42] D. Susser, B. Roessler, and H. Nissenbaum, “Online manipulation: Hidden influences in a digital world,” *Geo. L. Tech. Rev.*, vol. 4, p. 1, 2019.
- [43] V. P. Ta, R. L. Boyd, S. Seraj, A. Keller, C. Griffith, A. Loggarakis, and L. Medema, “An inclusive, real-world investigation of persuasion in language and verbal behavior,” *Journal of Computational Social Science*, vol. 5, no. 1, pp. 883–903, 2022.
- [44] Y. R. Tausczik and J. W. Pennebaker, “The psychological meaning of words: LIWC and computerized text analysis methods,” *Journal of Language and Social Psychology*, vol. 29, no. 1, pp. 24–54, 2010.
- [45] M. Van Haastrecht, B. Yigit Ozkan, M. Brinkhuis, and M. Spruit, “Respite for SMEs: A systematic review of socio-technical cybersecurity metrics,” *Applied Sciences*, vol. 11, no. 15, p. 6909, 2021.
- [46] M. T. Whitty and T. Buchanan, “The online dating romance scam: The psychological impact on victims—both financial and non-financial,” *Criminology & Criminal Justice*, vol. 16, no. 2, pp. 176–194, 2016.
- [47] M. T. Whitty, “Is there a scam for everyone? psychologically profiling cyberscam victims,” *European Journal on Criminal Policy and Research*, vol. 26, no. 3, pp. 399–409, 2020.
- [48] A. A. Williams, A. Lewis-Parks, and W. Wolf, “Do demographic, psychological, and financial characteristics increase the likelihood to be victims of credit card fraud?” *Journal of Personal Finance*, vol. 22, no. 2, 2023.
- [49] J. Wright, “‘many people are saying...’: Applying the lessons of naïve skepticism to the fight against fake news and other ‘total bullshit’,” *Postdigital Science and Education*, vol. 2, no. 1, pp. 113–131, 2020.
- [50] S. Zong, A. Ritter, G. Mueller, and E. Wright, “Analyzing the perceived severity of cybersecurity threats reported on social media,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 1380–1390.



Table 4: FEATURE EXTRACTION FOR INDIVIDUAL VULNERABILITY INDEX (IVI) FROM THE iPoll DATASET

Related Questions	Response and Score Mapping
<b>Lack of Awareness <math>A_{1,k}</math></b>	
Q7. Generally, how concerned, if at all, are you that you and/or a family member may fall victim to a scam?	Very concerned: 0, Somewhat concerned: 1, Not too concerned: 2, Not at all concerned: 3, DON'T KNOW/ SKIPPED ON WEB/REFUSED: ignore
Q8. How familiar were you with online romance scams; Q14: Grandparent scams? Q21: Government impostor scams; Q28 Census scams	Very: 0, Somewhat: 1, A little: 2, Not at all: 3, DON'T KNOW/ SKIPPED ON WEB/REFUSED: ignore
<b>Lack Knowledge of Protect Measure <math>A_{2,k}</math></b>	
Q37. Caller ID is a reliable way to know where a call is coming from?	True: 0; False: 5, Not sure: 3, SKIPPED ON WEB/REFUSED: ignore
Q38. When surfing the internet, it is always safe to interact with a website as long as the website has a locked box icon that indicates it is HTTPS secured.	Same as above
Q39. The IRS can call you about back taxes that you may owe without sending you a written notice first.	True: 5; False: 0, Not sure: 2, SKIPPED ON WEB/REFUSED: ignore
Q40. The Social Security Administration will contact you directly, either by phone or email, if there is a problem with your Social Security benefits.	Same as above
<b>Frequency of behaviors increasing the risk of attack <math>B_{1,k}</math></b>	
Q1. Not including time that you spend participating in online surveys, how often do you typically go online or access the Internet, including sending or receiving email?	Daily: 5, Several times a week: 4, several times a month: 3, Once a month: 2, Less than once a month: 1, Never: 0
Q2. How often, if at all, do you use the Internet to do the following activities	Daily: 5, Several times a week: 4, several times a month: 3, Once a month: 2, Less than once a month: 1, Never: 0
Q3. Have you ever done any of the following to meet potential dates or romantic partners at any point or time in your life?	Yes: 5, No: 0, Not sure: 3
<b>Trust level related to attack <math>P_{1,k}</math></b>	
Q6. How well do the following statements describe you?(Overall, I expect more good things to happen to me than bad; I sympathize with others' feelings; I am a trusting person; Overall, I am pleased with my life.)	very well: 5, Somewhat: 3, Not at all: 0
Q6. How well do the following statements describe you?(I find it difficult to get emotionally close to others; I worry a lot.)	very well: 0, Somewhat: 3, Not at all: 5
<b>Risk perception and impulsivity <math>P_{2,k}</math></b>	
Q5. Have you ever developed a romantic relationship with someone that you have never met in person?	Yes: 5, No: 0, DON'T KNOW/ SKIPPED ON WEB/REFUSED: ignore
Q6. How well do the following statements describe you? (I tend to get involved in things that I later wish I could get out of; I tend to make up my mind quickly; I feel uneasy in social settings	Very well: 5, Somewhat: 3, Not at all: 0
<b>Past Experiences <math>E_{1,k}</math></b>	
Q4. Thinking of the dates or romantic partners that you have met first online, have any of them ever done the following? (Lied about themselves, ask for money, etc.)	Yes: 5, No: 0, Not sure: 3
Q9. To the best of your knowledge, have you ever been a target of a romance scam; Q15, Grandparent scam; Q22. Government impostor scams; Q29. Census scams; Q35. Identify theft	Yes: 5, No: 0, Not sure: 3, SKIPPED ON WEB/REFUSED: ignore
Q12. To the best of your knowledge, has anyone you know ever been a target of a romance scam? Q19: grandparent scam. Q26. government impostor scams. Q33. Census scams	Yes: 0, No: 5, Not sure: 3, SKIPPED ON WEB/REFUSED: ignore
Q13. Did the person lose any money or suffer other financial losses due to the romance scam? Q20: grandparent scam. Q27. government impostor scams. Q34. Census scams	Same as above
Q36. Approximately, when did you experience identity theft?	less than a year: 1, 1-2 year: 2, 3-5 year: 3, 5-9 year: 4, more than 10 years: 5, don't know/skipped/refused: ignore
<b>Responses to past incidents <math>E_{2,k}</math></b>	
Q10. Have you ever lost money or suffered other financial losses due to a romance scam? Q17: grandparent scam. Q24. government impostor scams; Q31. Census scams	Yes: 5, No: 0, DON'T KNOW/ SKIPPED ON WEB/REFUSED: ignore
Q11. Have you ever experienced any health problems or emotional distress due to a romance scam? Q18: grandparent scam. Q25. government impostor scams; Q32. Census scams	Yes, health problems only: 3, Yes, emotional distress only: 3, both: 5, no: 0, DON'T KNOW/ SKIPPED ON WEB/REFUSED: ignore

Table 5: FEATURE EXTRACTION FOR THE ATTACK SEVERITY INDEX (ASI) FROM THE IPOLL DATASET.

Frequency Factor	
Q9. "To the best of your knowledge, have you ever been a target of a romance scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q12. "To the best of your knowledge, has anyone you know ever been a target of a romance scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q15. "To the best of your knowledge, have you ever been a target of a grandparent scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q19. "To the best of your knowledge, has anyone you know ever been a target of a grandparent scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q22. "To the best of your knowledge, have you ever been a target of a government impostor scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q26. "To the best of your knowledge, has anyone you know ever been a target of a government impostor scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q29. "To the best of your knowledge, have you ever been a target of a Census scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q33. "To the best of your knowledge, has anyone you know ever been a target of a Census scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Consequence Factor	
Q10. "Have you ever lost money or suffered other financial losses due to a romance scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q11. "Have you ever experienced any health problems or emotional distress due to a romance scam?"	Yes, health only: 4; Yes, emotional distress only: 4; Yes, both: 5; No: 0, SKIPPED/REFUSED: ignore
Q13. "Did the person lose any money or suffer other financial losses due to the romance scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q17. "Have you ever lost money or suffered other financial losses due to a grandparent scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q18. "Have you ever experienced any health problems or emotional distress due to a grandparent scam?"	Yes, health only: 4; Yes, emotional distress only: 4; Yes, both: 5; No: 0, SKIPPED/REFUSED: ignore
Q20. "Did the person lose any money or suffer other financial losses due to the grandparent scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q24. "Have you ever lost money or suffered other financial losses due to a government impostor scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q25. "Have you ever experienced any health problems or emotional distress due to a government impostor scam?"	Yes, health only: 4; Yes, emotional distress only: 4; Yes, both: 5; No: 0, SKIPPED/REFUSED: ignore
Q27. "Did the person lose any money or suffer other financial losses due to the government impostor scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q31. "Have you ever lost money or suffered other financial losses due to a Census scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Q32. "Have you ever experienced any health problems or emotional distress due to a Census scam?"	Yes, health only: 4; Yes, emotional distress only: 4; Yes, both: 5; No: 0, SKIPPED/REFUSED: ignore
Q34. "Did the person lose any money or suffer other financial losses due to the Census scam?"	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore
Sophistication Factor	
Q10, Q13, Q17, Q20, Q24, Q27, Q31, Q34	Yes: 5, No: 0, Not sure: 1, SKIPPED/REFUSED: ignore

Table 6: STATE-WISE SUMMARY OF MEAN IVI, ASI, SCVI, AND CONFIDENCE INTERVALS

State	Sample Size	Mean IVI	Mean ASI	Mean SCVI	CI Lower	CI Upper
Alabama	21	2.2017	1.1519	1.7487	1.2129	2.2844
Alaska	2	2.6621	2.5000	2.6511	0.4127	4.8894
Arizona	52	2.4559	1.3683	1.9859	1.6792	2.2927
Arkansas	14	2.2154	1.5321	1.8738	1.4236	2.3240
California	197	2.3034	1.4641	1.9769	1.8126	2.1411
Colorado	43	2.3126	1.2260	1.9192	1.5396	2.2989
Connecticut	648	2.2852	1.0578	1.7606	1.6762	1.8451
Delaware	8	2.2472	0.4125	1.3298	1.2127	1.4470
District of Columbia	8	1.9101	0.4538	1.1819	0.8204	1.5435
Florida	105	2.3152	1.4808	2.0194	1.7806	2.2583
Georgia	33	2.3542	1.5691	2.1189	1.6624	2.5753
Hawaii	5	2.4058	1.1220	1.7639	1.0852	2.4426
Idaho	13	2.1801	1.1246	1.8311	1.0335	2.6286
Illinois	67	2.2632	1.3351	1.8928	1.6040	2.1815
Indiana	38	2.3518	1.2792	1.8571	1.5881	2.1261
Iowa	24	2.1667	1.0788	1.7283	1.2274	2.2291
Kansas	17	2.4039	1.4394	1.9824	1.5044	2.4603
Kentucky	12	2.1597	1.1825	1.6711	1.1725	2.1696
Louisiana	16	2.1480	1.2581	1.7031	1.2765	2.1297
Maine	10	2.3044	1.3910	1.9607	1.2731	2.6483
Maryland	21	2.2142	1.1205	1.7728	1.2382	2.3074
Massachusetts	39	2.3466	0.9431	1.7233	1.4070	2.0396
Michigan	51	2.1960	1.3165	1.8337	1.5217	2.1457
Minnesota	25	2.4252	1.4620	2.1542	1.6321	2.6763
Mississippi	5	2.3783	2.0000	2.2781	0.9182	3.6381
Missouri	41	2.3160	1.4295	1.9827	1.6268	2.3387
Montana	10	2.1225	1.4240	1.8202	1.0690	2.5715
Nebraska	26	2.2278	1.0808	1.6977	1.3492	2.0463
Nevada	11	2.3484	2.3745	2.4792	1.7692	3.1891
New Hampshire	3	2.1882	1.4300	1.8091	1.6056	2.0126
New Jersey	31	2.1925	1.4452	1.9661	1.5002	2.4320
New Mexico	14	2.2762	1.5393	2.0750	1.3655	2.7845
New York	71	2.1837	1.0558	1.6327	1.4244	1.8410
North Carolina	55	2.2854	1.4275	1.9378	1.6379	2.2377
North Dakota	3	2.4277	1.5400	1.9838	0.9936	2.9741
Ohio	57	2.3874	1.2839	1.9111	1.5969	2.2253
Oklahoma	658	2.3456	1.1328	1.8324	1.7468	1.9179
Oregon	20	2.2713	0.8440	1.6130	1.2306	1.9954
Pennsylvania	666	2.2713	1.0643	1.7417	1.6594	1.8240
Rhode Island	6	2.3661	2.3817	2.7535	1.2373	4.2697
South Carolina	24	2.2538	1.2692	1.8270	1.3790	2.2750
South Dakota	5	2.3018	0.4620	1.3819	1.2523	1.5115
Tennessee	29	2.4633	1.2490	2.0675	1.5640	2.5710
Texas	104	2.2039	1.1030	1.7012	1.5040	1.8984
Utah	12	2.5477	1.5483	2.1577	1.4140	2.9014
Vermont	465	2.3695	1.1945	1.8394	1.7480	1.9308
Virginia	36	2.2987	0.9575	1.7331	1.3814	2.0848
Washington	671	2.3105	1.2937	1.8845	1.7982	1.9708
West Virginia	10	2.1353	0.5610	1.3481	1.0176	1.6787
Wisconsin	48	2.2806	0.6758	1.5271	1.2690	1.7851
Wyoming	5	2.5281	1.0560	1.7921	1.0546	2.5296