Anticipating Personal Cyber Insurance Disputes: A US/UK User Study

Temima Hrle
University of Edinburgh
temima.hrle@ed.ac.uk

Yangheran Piao
University of Edinburgh
lawrence.piao@ed.ac.uk

Daniel Woods
University of Edinburgh
daniel.woods@ed.ac.uk

June 2, 2025

Abstract

Personal cyber insurance has emerged in recent years to cover various digital risks associated with online fraud, security incidents, privacy violations, and even cyberbullying. This improves risk management options for individuals who understand the underlying risks and insurance coverage. However, misinformed buyers risk having claims denied, sometimes taking riskier decisions under the false belief that insurance covers the consequences. Cyber insurance is especially likely to create misunderstandings because both the insurance product and underlying risks are new to the customer.

This paper explores the potential for future disputes by examining the gap between public understanding of cyber perils and the definitions found in insurance policies. We use a survey instrument to collect 3,234 definitions of the major perils under personal cyber insurance. Participants defined each harm in their own words, which were qualitatively coded. The codebook captured structural elements, victim-adversary relationships, actions, motivations, technical specifics, and impacts on victims. We then mapped participant definitions to policy language to identify areas where public perceptions diverge from the coverage offered by cyber insurance policies. Our results show that participants correctly identify actions associated with cyber harms to varying degrees, and that education level and income have little effect on how correctly harms are defined. Our research has important policy implications that reiterate the need for clearer and more accessible insurance policy language.

1 Introduction

Since the late 2010s, personal cyber insurance has emerged to cover digital risks including online fraud, ransomware, and cyberbullying [1, 2]. This coverage is much broader than personal identity insurance—available since the early 2000s— that only covers losses arising out

of identity theft [3]. Both personal products have parallels with corporate cyber insurance, which has been available since the late 1990s [4, 5]. The majority of research has focused on the corporate product [6–9], with a recent review noting a "paucity" of research into personal cyber insurance [10].

This matters because personal products require stronger consumer protections, and new products require yet greater oversight. Cyber insurance was historically bought by large organizations with in-house legal expertise and IT teams who work with independent brokers to understand and negotiate insurance contracts [5]. Despite this expertise, companies ended up in disputes over coverage for cyber incidents over issues like whether electronic data constitutes tangible property [11].

Meanwhile, individuals typically have less legal expertise, and are less likely to be supported by an independent broker [12]. The potential for misunderstandings can be seen in the reality that consumers struggle to understand even well-established products like health insurance [13], let alone new products covering harms that did not exist until a few decades ago.

Misinterpretations may lead policyholders to falsely believe they are covered for incidents, leading to at least three types of problems. First, the individual is unexpectedly liable for a large financial loss, a negative shock that can impact the family of the victim's well-being. Second, this undermines trust in the insurance product, damaging the industry and reducing options to the consumer. Third, this intensifies moral hazard as the insured believes they have even more coverage, taking even riskier decisions at the margin.

To explore the scope of the misunderstanding problem, this study investigates how the public defines cyber harms and whether their conceptualizations align with the definitions and coverages provided in personal cyber insurance policies. We ask the following research questions:

RQ1 How do individuals define different types of cyber harms ?

RQ2 Do participant definitions align with coverage in cyber insurance policies ?

RQ3 How do demographic factors influence participants' definitions of cyber harms ?

Using a qualitative coding framework, we analyze 3,234 participant-provided definitions found in the six core perils covered by personal cyber insurance policies in the US and UK [14]. By mapping these definitions to existing cyber insurance coverages, this research highlights potential gaps in consumer understanding and identifies areas where policy language may contribute to misinterpretation.

Section 2 outlines the methodology adopted. Section 3 presents our findings across various metrics in our codebook. This section addresses the structure and specificity of definitions, and actions, motivations and impacts detailed by participants. Section 3.2 maps our definitions to cyber insurance coverages, to evaluate whether participants' identified harm characteristics are covered by policies. Section 4 discusses implications of low cyber harm literacy and how this is reflected in insurance policies.

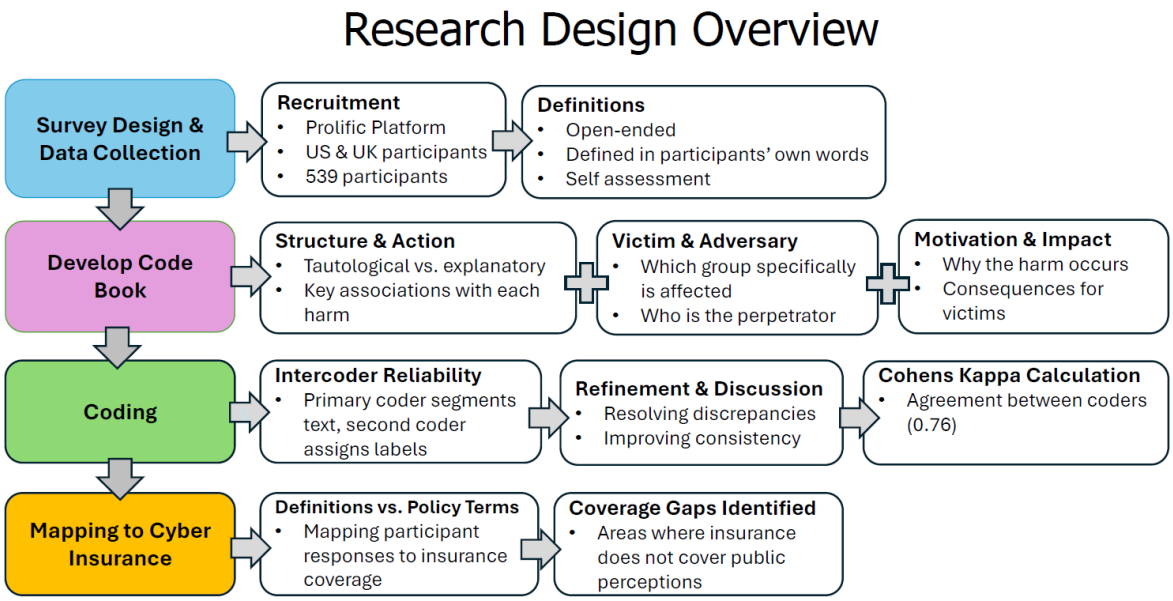


Figure 1: An overview of our research design.

2 Methodology

We conducted a qualitative analysis of responses to an online survey, which comprised 3,234 definitions across six cyber perils: cyber attack, cyber extortion, identity theft, data breach, cyberbullying, and online fraud. Figure 1 provides an overview of our research design. Section 2.1 details our data collection, including survey design, participant recruitment, and the demographics of our sample. Section 2.2 describes our approach to analyzing these definitions.

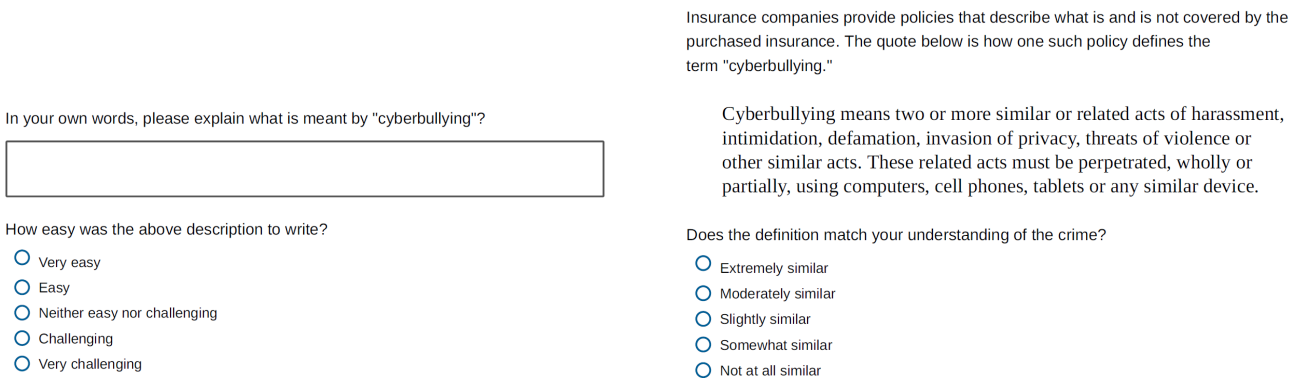
2.1 Data Collection

Our goal was to understand how participants understand the perils covered by personal cyber insurance. We extracted the most commonly covered perils from a study of personal cyber insurance in the US and the UK [14]. These incidents were used to design a survey instrument (see Section 2.1.1). We recruited participants via the Prolific crowd sourcing platform (see Section 2.1.2).

2.1.1 Survey Design

We first drafted a survey based on discussions with the core research team. The draft survey was circulated with colleagues working on human-centred security, refining it based on their feedback. We then piloted the survey on around 20 individuals recruited using the same approach to how we recruited the final sample (see Section 2.1.2). The responses appeared to be meaningful, despite the expectations of some colleagues that participants would not provide a detailed response to free-text questions.

The main tasks in the survey asked participants to define each of the cyber perils in their own words (see Figure 2), randomizing the ordering of perils each time. We decided for



(a) First Page (b) Subsequent Page

Figure 2: Format of survey questions asking participants to define individual harms and self-evaluate ease.

an open-ended task in order to access participants' understandings without priming them by, for example, presenting a multiple-choice list of possible definitions, which would not be available in the real-world when deciding whether to buy cyber insurance.

Following each definition, we asked participants to rate how easy or difficult it was to write that description, using a five-point scale; "very easy", "easy", "neither easy nor challenging", "challenging" or "very challenging". Following each definitions task, participants were then shown a sample insurance policy definition for the same harm. We then asked them to assess how closely it matched their own understanding of the term on a five-point scale: "extremely similar", "moderately similar", "somewhat similar", "slightly similar", or "not similar at all" (see Figure 2).

In addition to the main survey tasks, we collected demographic information and a light-weight security awareness scale, SA-6 [15]. We added two 'attention check' questions to the survey. We excluded responses from participants that failed either of the two attention check questions, although participants were still paid.

2.1.2 Participant Recruitment

Participants were recruited through Prolific, an online crowdsourcing platform designed for academic research. Prior work has demonstrated that Prolific produces higher quality data compared to other platforms like MTurk, particularly for studies on security and privacy perceptions [16]. Prolific has been shown to generalize well for questions about user experiences, perceptions, and beliefs. This provided a tolerable balance of convenience and reliability.

We recruited an even split of participants from the US and UK. These geographies match the study from which we extracted the cyber perils, which analyzed policies from the US and UK [2]. The survey was offered solely in English, a pragmatic choice given the research team lacked the resources to create the survey in other languages. Participants received a reward for their participation, provided through Prolific, with payment contingent on successful completion of the survey.

A full demographic breakdown is provided in the Appendix (see Table 3). The majority of participants were aged 30-39 years (33%), followed by 18-29 years (22%) and 40-49 years (18%). The majority of participants were white 82%. In terms of education, 40% of participants held a bachelor's degree, 15% had a master's degree, and 4% had a doctorate. High school graduates made up 14% of the sample. The gender distribution was balanced, with slightly more female (55%) than male (44%) participants. Participants reported a range of household incomes, with 16.3% in the highest income bracket ($100,000 or more) and 5% earning up to $25,000. Income in pounds ranged similarly, with 8% earning up to £20,000 and 6% earning £80,000 or more. Regarding employment status, the majority (53%) were employed full-time, with smaller proportions employed part-time (14%) or self-employed (10%).

2.2 Data Analysis

Section 2.2.1 describes our process for coding definitions. Section 2.2.2 explains how we used the codes to map definitions to coverage.

2.2.1 Coding Definitions

We developed a codebook via an iterative process, beginning with an open coding phase. This phase allowed us to identify patterns in how participants defined each harm, with the goal of systematically assessing whether their understandings would meet the criteria for coverage under typical cyber insurance policies (RQ2). The identified patterns formed the basis for our preliminary codes. This preliminary set of codes was then refined through several rounds of discussion and testing within the research team. The final codebook (see the Appendix) consisted of multiple categories of codes and subcodes, which were designed to assess how well participant definitions matched the criteria for insurance coverage.

The most expansive category in our codebook focused on the actions described in participant definitions, which we mapped to the six cyber perils. The analysis examined how well participant definitions aligned with the actions attributed to to each harm in the codebook. For instance, many participants accurately defined cyber extortion as involving blackmail or ransom, while others misunderstood identity theft, conflating it with data breach.[1]

We also included categories for types of victims and adversaries, making a distinction between the usage of first person in definitions, i.e. "when I am harassed online", indicating

[1]For instance, several participants captured cyber extortion by emphasizing ransom demands – one participant wrote, "cyber extortion is like a digital form of blackmail. It happens when hackers or cybercriminals threaten to expose sensitive or embarrassing information about individuals or organizations unless they pay a ransom," clearly referencing the coercive nature of blackmail. In contrast, identity theft was often misunderstood; for example, one respondent defined it as "when hackers steal personal information from a company's database," conflating identity theft with a data breach. The actions associated with data breaches were broken down into three categories - the first being the unlawful theft, access or use of data, the second being the unlawful disclosure of data, and thirdly, ambiguous data loss. Participants frequently failed to make the correct distinction between personal data being stolen or disclosed, and actively impersonating someone by pretending to be them, as correctly defined in an identity theft definition; "Using someone's identity to open credit cards, receive tax refunds, etc.".

that the participant was affected by the harm, the use of second person language, i.e. "you", and the distinction between generic third person references and entities as victims.

The motivation category captured the rationale behind committing the harm, such as financial gain, impersonation, blackmail or extortion, information gain, destructive or recklessness, and "other" motivations not captured by any of the aforementioned codes. Although not asked in the survey, any impact on victims as a result of a harm, identified by participants, was coded, i.e., "...online fraud can have serious consequences for the victim, including financial losses, damage to their credit score, and reputational harm." The impact section was broken down into two categories: direct and response costs, making it easier to compare impacts to cyber insurance coverages.

To test the reliability of the coding, the primary coder systematically broke down each participant definition into discrete segments, each reflecting an individual code within the codebook. A second coder assigned the appropriate codes to each segment. We discussed and resolved discrepancies, refining the codebook where necessary. We calculated Cohen's kappa to assess the agreement between coders and had a result of 0.76 across 172 units, indicating substantial agreement.

2.2.2 Mapping Definitions to Coverage

We mapped participants' definitions to cyber insurance coverage to identify the potential for future coverage disputes. This mapping was challenging given the task was open-ended, and many participants did not mention specific impacts. We considered two primary approaches to this mapping: (1) strict alignment criteria where only definitions that matched specific terms in insurance policies would be considered aligned with coverage; and (2) fuzzy alignment that allows for partial matches or conceptual overlaps in coverage.

We apply both approaches to our codebook from Section 2.2.1, rather than re-analyzing all 3,234 definitions. The action codes were mapped to the six perils, manually mapping any actions coded as miscellaneous.[2]

Impact codes were mapped to specific coverage costs where available, which was relatively rare. For example, no participants mentioned the costs of relocating to another school, or childcare and eldercare as a result of bullying, even though both can be covered by cyber insurance [2]. We manually filtered all 'miscellaneous impact' codes for individual cyber harms to see whether participants had mentioned a specific cyber insurance coverage.

We did not map definitions to broad or unspecified coverage, such as 'other,' 'other specified costs,' and 'other non-specified costs,' [2]. Instead we focused more specific coverage categories, such as "attorney fees and expenses", "lost wages", and "costs for credit reports" under identity theft or "educational expenses" and "relocation expenses" under cyberbullying. We first mapped the definition to the most likely insuring agreement. These were then manually filtered through to see whether individual definitions mention the specific coverage. This resulted in a low alignment rate, highlighting a significant limitation in understanding how users perceive the aftermath of such harms compared to what is covered by insurance.

[2]For certain harms such as cyber extortion, where the associated action is blackmail, and is also classified as a motivation in our codebook, both aspects were considered when mapping definitions to ransom payment coverage in cyber insurance policies. This approach is necessary because this coverage is typically triggered when an individual is blackmailed or extorted financially, meaning the action itself is also the impact.

Cyber Harm	Character Count in Definitions					Other Metrics	
	Mean	Median	Min	Max	Std Dev	Tautology Rate	Individual Victim
Cyber Attack	89.6	77	2	483	59.4	5.6%	80.1%
Cyber Extortion	89.4	77	3	544	64.5	6.3%	94.8%
Cyberbullying	81.4	64	10	964	72.9	20.6%	99.8%
Data Breach	89.6	76	8	584	61.4	9.8%	84.4%
Identity Theft	100.4	88	15	576	60.4	5.8%	100.0%
Online Fraud	89.3	76	6	507	60.1	14.5%	97.8%

Table 1: Character count statistics for definitions of each harm. Short minimum character counts are the result of nonsensical definitions such as "Na", "Idk" or "unsure". Other metrics such as tautology rates and the percentage of participants identifying individuals as opposed to entities as victims are also provided.

2.3 Responsible Research

This research project received ethical approval from our institution. Prior to starting the survey, participants were required to complete a consent form, upon which they were redirected to the questions. Participants had the right to withdraw from the study at any time, although none used it.

In terms of compensation, the reward was set based on an estimated completion time of 20 minutes multiplied by the National Living Wage in the UK. This is higher than most state-level minimum wages in the US, although is likely lower than affluent urban areas.

3 Results

Section 3.1 focuses on how participants construct definitions (**RQ1**). Section 3.2 turns to the alignment with personal cyber insurance coverage (**RQ2**).

3.1 Definitions (RQ1)

We start by discussing the structure and specificity of definitions, before asking whether the semantic meaning matches "correct" definition according to the insurance industry.

3.1.1 Structure and Specificity of Definitions

Information Content Participants provided varying levels of detail across different cyber harms. The crude character counts in Table 1 suggest that definitions of cyberbullying have the least information content, identity theft have the most, and the other four perils all have similar information content.

This partially aligns with the perspective provided by the tautology rate. A representative example of a tautological definition of cyberbullying is "bullying online". In contrast, identity theft exhibited a low tautological rate (6%), and had the highest explanatory rate (94%), with the longest average character count. It is the only harm that all participants attempted to define (even though it was most frequently incorrectly defined).

Cyber Harm	Correct Broad	Correct Narrow	Mention of Impact (%)	Most Common Impact	Most Common Motivation
Cyber Attack	52.7%	41.9%	2.1%	Repu. Damage	Info. Gain (25.4%)
Cyber Extortion	69%	60.9%	0%	Financial Loss	Fin. Gain (48%)
Identity Theft	49%	35.1%	0.8%	Repu. Damage	Fin. Gain (28.2%)
Data Breach	95.7%	83.5%	0.2%	Misc. Impact	Info. Gain (3.3%)
Cyberbullying	99.8%	86.7%	2.6%	Mental Anguish	Emotional Harm
Online Fraud	58.3%	49.7%	0.8%	Financial Loss	Deception (36.2%)

Table 2: Table showing the correct action attribution under broad and narrow correctness, along with the most common impacts and motivations identified in definitions.

Longer definitions suggest that participants perceive the peril as a more complex or multifaceted harm, prompting more elaborate definitions compared to other perils. Meanwhile, shorter descriptions such as those for cyberbullying imply that participants view the harm as self-explanatory, hence requiring less detail to define. This interpretation is supported by the thematic density across definitions.

Across all perils, cyberbullying exhibited the highest concentration of single-theme definitions. In contrast, cyber attack and online fraud demonstrated greater thematic diversity. Definitions of cyber attacks included multiple themes, such as 'hacking' (43%) and 'denial of service' (10%), both of which are actions correctly associated with cyber attacks.

Actors We also identified patterns in how participants associate perils with victims and adversaries (see Table 1). Unsurprisingly, most participants associate the victims of cyber perils with individuals rather than entities. For harms such as identity theft and cyberbullying, individuals were almost entirely identified as the primary victims (100% and 99.8% respectively). In contrast, data breach and cyber attack definitions were the most likely to include organizations as victims (16% and 20% respectively). This may stem from the nature of these incidents, where organizational data breaches often impact large organizations [17]. Definitions of cyber extortion and online fraud rarely invoked organizaitons as victims (5% and 2% respectively), even though ransomware incidents are a common type of cyber incident.

Perpetrators varied significantly across definitions. The security perils (data breach, cyber attack, and cyber extortion) were most likely to mention a specific adversary like "hackers," "scammers," or "criminals" (16%, 13% and 7% respectively). This specificity may result from media narratives that frequently link data breaches to targeted attacks by organized groups. In contrast, only 1% of cyberbullying definitions referred to specific perpetrators like a "bully" or "harasser". This could result from a perception that cyberbullying is a societal issue rather than actions of a separate set of actors. Definitions that mentioned specific perpetrators were more likely to include motivations for harms.

3.1.2 Actions, Motivations and Impacts

We introduce the notion of "correctness" to describe actions, distinguishing between broad and narrow correctness. The broad notion require a participant to attribute the correct action to the harm, regardless of whether additional actions are attributed. The narrow notion

requires participants to not include any action that would qualify as correct under another peril. For example, defining cyber attack as "a hacking event such as DDOS, ransomware or data breach" would be broadly correct, but would fail the narrow correctness criterion as this definition includes an action that is correct for data breach. This helps to track whether definitions have a concise conceptual meaning, or whether they sprawl across categories.

Table 2 shows cyberbullying and identity theft are, respectively, the most correctly and incorrectly defined harms under both the broad and narrow interpretations. Cyberbullying, which had the highest accuracy, also had the highest tautology rate (21%), despite this, these responses were technically accurate under a narrow definition criterion. This suggests that while cyberbullying is an intuitive harm, it lacks depth in how participants describe its nuances. Table 2 also highlights the most common motivations and impacts for each peril. We now dive into the specific perils.

Cyberbullying was the least diverse, and most well understood harm in terms of action. Most definitions (74%) did not specify behaviors other than "bullying", suggesting that participants believe that parallels with non-digital bullying do not require clarification. Cyberbullying was slightly more likely to be viewed as occurring in private interactions between users such as through direct messages (12%) as opposed to public forums, such as Facebook posts and Tweets (8%). Few definitions (3%) recognized that it could occur in both private and public spaces. Harassment emerged as a prominent theme, appearing in a quarter of definitions. Only a subset of participants specified its repetitive nature, indicating varied interpretations of harassment within the context of cyberbullying.

Cyberbullying exhibited the lowest rate of motivation mentions, with only 5% of participants identifying any motive. These mostly comprised vague motivations, with specific goals almost entirely absent. In terms of the impact on victims, participants overwhelmingly emphasized cyberbullying's psychological toll. Mental anguish accounted for the vast majority (79%) of impacts. Miscellaneous effects, such as "profound distress", and reputation damage made up 14% and 7% of the impacts respectively.

Identity theft had the lowest narrow definition rate (35%), which means it was the most frequently misattributed peril. Most participants associated identity theft with data breaches, with a slim majority not identifying the impersonation action at all. This finding suggests that identity theft is frequently confused with data breach, even though the reverse does not occur. No participants conflated data breach with impersonation.

Identity theft is an example where the narrow definition is not appropriate because defining identity theft as stealing data *and* impersonating the victim to take out a loan, which was done by 13% of participants, is not incorrect. However, only defining it as stealing data is incorrect as the defining action is impersonation.

Participants most frequently attributed the motivation for identity theft to financial gain (28%), followed by impersonation (13%). It had the highest proportion of unspecified or "other" motivations (31%), with motivations often described in vague terms such as the perpetrator gaining "something" for "malicious purposes" or to "commit crimes". A tiny share (1%) of identity theft definitions included impact codes.

Cyber Attack was correctly identified as involving actions like hacking (43%) and denial of service (DoS) (10%). However, 37% of responses fell into the miscellaneous category, where participants described actions they associated with the harm in ways that do not fit the predefined subcodes. This was common in instances where a motivation for the

harm was provided in lieu of the crime's defining action, such as "When someone performs a malicious act over the internet intended to gain confidential information or just cause chaos." In such examples the motivation, as opposed to the action is the focal point. The majority of definitions in this category provided overly vague actions such as "It [cyber attack] is when you do anything online to harm another person" or "Crime over the internet".

The considerable range of definitions for cyber attack may be as a result of participants viewing the harm as encompassing a wide range of incidents. The most commonly cited motivation was information gain (25%), followed by destructive intent (11%). Financial gain (4%) and blackmail (3%) were less frequently mentioned, while impersonation (1%) was almost entirely absent. Furthermore, 12% of participants provided vague motivations.

Cyber attack definitions infrequently (2%) mentioned an impact, of which half were classified as miscellaneous impacts. These included damage to systems, website crashes, and the spread of viruses. This reflects participants' perception of cyber attacks as technical disruptions with broad and immediate consequences. Reputational damage was the second most frequently mentioned impact (27%), underscoring participants' recognition of the harm's ability to undermine an organization's public image. Additionally, legal costs (9%) and mental anguish (9%) were cited less often, suggesting a secondary focus on psychological and financial consequences.

Cyber extortion emerged as one of the most accurately defined harms, with 69% of participants correctly identifying blackmail or threatening as the core action. A small minority (6%) of participants linked the harm to data theft, which suggests that stolen data is a prerequisite for extortion. This is an example where the narrow correctness criteria is perhaps more meaningful, given the cyber action is associated with the extortion demand, regardless of whether data is stolen, encrypted or reputation is threatened. Notably, around 5% of participants defined cyber extortion using the cyber attack action.

Almost half of participants (48%) cited financial gain as the primary motivation for cyber extortion, accounting for 87% of those who provided any motivation. This strong alignment between financial incentives and the act of blackmailing or threatening someone, correctly identified by 69% of participants, emphasizes its straightforward economic agenda. Cyber extortion definitions did not address the impact on victims. Only 2% of participants mentioned motivations like blackmail or information gain, even though threats to leak sensitive images is often part of intimate partner violence [18].

Data breach definitions overwhelmingly described data theft as the core action 64%. An equal number of participants (16%) defined data breach as the unlawful disclosure of data, and an ambiguous data leak. Data breaches could be made up of any of the aforementioned actions in our codebook.

Motivations were rarely included in data breach definitions, with only 3% associating the harm with information gain. Financial gain (1%) and blackmail/extortion (0.4%) were infrequently cited. Despite the high specificity of adversary mentions (16%), participants mentioned the impacts of data breaches even less frequently (0.2%), primarily as miscellaneous consequences. Participants often portrayed data breaches as technical events rather than deliberate acts driven by specific goals.

Online fraud was correctly defined as trickery or deceit by over half of the participants (58%). Impersonation followed as the second most common action (27%), which suggests participants consider identity theft to be a subset of online fraud. These definitions should

not be considered incorrect, but they potentially underestimate the scope of coverage.

Unsurprisingly, financial gain was the most frequently cited motivation for online fraud (36%), emphasizing participants' understanding of the harm as an economically driven crime. Some participants (10%) cited information gain, reflecting some recognition of fraud as a means to acquire sensitive data beyond immediate financial rewards. A tiny share (1%) of online fraud definitions included addressed the impact, focusing primarily on financial loss.

3.2 Insurance Coverage (RQ2)

We analyzed how closely participants' definitions of cyber harms aligned with the categories typically covered by cyber insurance policies. This process seeks to identify areas where user expectations align with, or diverge from actual insurance coverages. It provides insight into whether participants define harms in congruency with insurance coverages, in addition to any gaps that policies may have missed, as reflected by definitions.

Identity Theft None of the definitions identified costs that could be mapped to identity theft coverages, which are fine-gained costs for credit reports, credit monitoring, or re-filing application costs as a result of rejection due to the attack [2]. There were only three definitions under our 'miscellaneous' code in the impact category of our codebook, none of which could be mapped. Financial harm was mentioned as the impact in all three definitions, which is not covered by cyber insurance in most cases. This lack of alignment is perhaps unsurprising given many participants misunderstood the action that constitutes identity theft.

Cyber Attack The coverages related to cyber attacks are cyber disruption services, data recovery and system restoration [14]. Participants did not consider the necessity of disruption services after an attack, which is the cost of hiring an expert to restore the system back to the state it was before the attack. They tended to focus on immediate outcomes (i.e., denial of service or hacking) rather than the broader consequences that might require ongoing support, such as disruption services or data recovery efforts.

Since the majority of participants accurately associated cyber attacks with hacking events or denial of service, all correct definitions shared the understanding that the system is compromised. Participants provided detailed examples, such as: "The consequences can be that the website crashes, is unable to process genuine users, or even that sensitive data is stolen," "A website or computer system crashing because it has been DDOSed," or "... which resulted in loss of service or damage to the system." These examples demonstrate complete alignment with system restoration coverage in insurance policies, as they reflect an understanding of the harm's impact in terms of system compromise and the need for restoration. Additionally, 17% of participants expressed an understanding of the need for data recovery, citing scenarios where data is lost as a result of the incident.

Cyber Extortion All definitions where blackmail/extortion was identified as the motivation (100%) mapped directly to ransom payment coverage. This indicates participants' strong understanding of cyber extortion as a financially motivated harm. However, only 77% of definitions identifying blackmail/extortion as the action could be mapped to ransom payment coverage. This discrepancy arises because participants often described cyber extortion in generic terms, such as "online blackmail," without naming the ransom payment. This could represent a misunderstanding as the ransom demand could constitute a non-monetary

request, such as for intimate images.

In contrast, no alignment was observed for broader or vague insurance coverage categories, such as "unspecified professional assistance," "other non-specified costs," and "other specified costs." These categories, inherently ambiguous in policy language, did not correspond to participant definitions, which focused primarily on explicit actions like blackmailing or threatening. This disconnect highlights a challenge in bridging participant expectations with generalized policy terms, as participants describe specific, immediate actions and outcomes rather than encompassing broader or unspecified impacts.

Cyberbullying The definitions demonstrated varied alignment with cyberbullying coverage, with some categories achieving full alignment while others were completely unaddressed by participants. The strongest alignment occurred for mental health services, where all definitions mentioning mental anguish as an impact were directly mapped to this coverage. This reflects participants' perception of cyberbullying as a harm that primarily affects psychological well-being, aligning closely with the mental health services offered under insurance policies.

In contrast, there was no alignment for several other coverage categories, such as educational expenses, relocation expenses, salary lost, legal expenses, childcare or eldercare expenses, and the purchase of support software. These categories, while relevant in some insurance contexts, were entirely absent from participants' definitions. This suggests a limited public understanding of the broader or secondary impacts of cyberbullying, such as the financial or logistical burdens that may arise in severe cases.

Data Breach The alignment between data breach definitions and insurance coverages revealed significant gaps, reflecting participants' narrow framing of the harm as predominantly technical, with limited consideration for broader consequences or recovery processes. None of the reviewed definitions aligned with coverage categories such as legal professional assistance, IT professional assistance, notification costs, or services like "services to affected individuals". Participants did not include generalized impacts or the secondary consequences that such categories aim to address, suggesting that definitions of data breaches remain narrowly focused on the immediate action of data theft rather than the subsequent recovery efforts.

Online Fraud Online fraud demonstrated complete alignment between the financial fraud insurance coverage and participant definitions where financial gain was identified as the motivation. Participants consistently described online fraud as an economically driven harm, emphasizing its direct financial intent. For example, many participants framed the harm in terms of scams or deceptive practices aimed at monetary exploitation.

Alignment was significantly lower (20%) when mapping the direct financial fraud coverage to participant definitions categorized under miscellaneous impacts. These definitions often referred to vague or indirect consequences of fraud, such as inconvenience or stress, which did not explicitly connect to the direct financial loss resulting from the attack. This reflects a surprising discrepancy—participants did anticipate stress and extra work associated with online fraud incidents, however insurers did not offer coverage for this stress or the inconvenience. The opposite is true of identity theft, which covers inconvenience via lost wage and care costs, and cyberbullying, which covers stress via mental health support costs.

3.3 Correlates of Understanding (RQ3)

To better design interventions, we explored how individual factors correlated with understanding of the perils. Although we collected demographics data like gender and race, there was no clear rationale for how these factors would influence understanding [19]. Instead we focused on the following hypotheses, tested using a Chi-squared test[3]:

H1: Participants with higher income will more frequently define perils correctly.

H2: More educated participants will more frequently define perils correctly.

H3: More educated participants will more frequently provide explanatory definitions.

These hypotheses are motivated by the reality that the survey task of defining cyber perils is knowledge based. Knowledge workers may perform better because they are better at constructing definitions, and not because of any understanding of the peril. However, the chi-squared test showed there was no significant association between correctly defining perils and income (**H1**: $\chi^2(df = 4, N = 3234) = 7.8, p = 0.167$). Similarly, there was no significant association between education level and correctly defining perils (**H2**: $\chi^2(df = 5, N = 3234) = 3.4, p = 0.636$), or providing explanatory definitions (**H3**: $\chi^2(df = 5, N = 3234) = 10.5, p = 0.061$). The research team was surprised by these results, which suggest performance on the definition task was independent of education level.

One explanation for no relationship in H1–3 could be noise in the task, response, or our criteria for correctness. To test this, we used the following hypotheses:

H4: Participants' who believe a peril was easier to define will more frequently define perils correctly.

H5: Participants' who believe their definition was similar to the insurance definition will more frequently define perils correctly.

Based on our pilot, we believed that participants were meaningfully carrying out the task. Indeed, there was a significant association between correctness of definitions and both: perceived ease ((**H4**: $\chi^2(df = 4, N = 3234) = 105.3, p < 0.001$) and perceived similarity to the insurance definition ((**H5**: $\chi^2(df = 4, N = 3234) = 81.3, p < 0.001$). This supports our belief that our research design is meaningful.

Finally, we wanted to test whether security awareness (according to SA-6 [15]) has a relationship to the correctness of definitions. We made the following hypotheses:

H6: Participants with higher security awareness will more frequently define perils correctly.

H7: Participants with higher security awareness will more frequently provide explanatory definitions.

[3]We used each definition task as a data point, assigning a score of 0 if incorrect/tautological and 1 if correct/explanatory. Consequently, each participant provided six data points.

We expected cybersecurity awareness, which involves understanding cyber incidents, to improve the accuracy of definitions. There was no significant association between higher security awareness and correctly defining perils (**H6**: $\chi^2(df = 23, N = 3234) = 24.6, p = 0.369$). However, there was a significant association between higher security awareness and providing an explanatory definition (**H7**: $\chi^2(df = 23, N = 3234) = 41.7, p < 0.01$). This suggests that participants with higher security awareness are less likely to use tautologies, but no more likely to correctly define the peril.

4 Discussion

The misalignment between public perception of cyber harms and insurance coverage creates trust and usability challenges. A key question is whether these gaps reflect a failure of insurance policies to adequately protect consumers, or if consumers themselves lack the necessary understanding of these terms and thereby insurance coverages. The answer motivates different policy responses.

4.1 Improving Insurance Policies

From a legal theory perspective, the challenges can be framed through the lenses of consumer protection law. Within these frameworks, the consumer's right to understandability and clarity becomes paramount. If consumers purchase coverage for complex cyber risks, insurers must provide clear and comprehensible terms that reflect the consumer's risk landscape. Legal principles require that terms be presented in ways that match the reasonable expectations of an average consumer, which is essential for enforcing a valid contract.

Insurers can address this issue by refining policy language to make it more accessible and aligned with consumer understanding. This can be done through the standardization of definitions of cyber harms to align with common public understanding. Moreover, regulators may need to enforce stronger guidelines on how personal cyber insurance products are marketed, ensuring clarity and reducing the risk of consumer confusion. The harm that raises the greatest concern is identity theft, which less than half of participants defined correctly, even using a broad reading, often conflating it with data breach. This presents a real-world problem given many home insurance policies extend digital coverage only for identity theft [2], and there are standalone identity theft only products [3]. A simple solution is for insurers to always offer data breach coverage alongside identity theft. The consequence would be that consumers are only confused as to which insuring agreement is triggered, but that this confusion will not result in a rejected claim, providing an extra layer of security.

There is also an interesting question about the right level of specificity across harms. Perils like cyber attack and online fraud provide broad coverage, which is good, but they also create confusion. Insurers could, for example, include perils for both hacking and Denial of Service instead of including both under the umbrella term, cyber attack. This would likely be clearer for consumers, who mentioned that cyber attack is confusingly broad. However, distinction also brings complexity. Many consumers would prefer if home insurance policies included an umbrella term like "physical disaster" instead of attaching different coverage to individual perils like fire, flood, wind and so on.

It is additionally necessary to address the gaps in coverage between cyber harms, such as by creating a distinction, for example, between traditional interpretations of cyberbullying and revenge porn. Cyberbullying definitions frequently provide such an example: "cyberbullying happens online and can take various different routes such as doxxing, personal harassment, revenge porn, personal info being leaked, kids bullying etc."

Revenge porn is typically excluded from the scope of insurance, the omission of which highlights a growing limitation within the cyber insurance market—particularly the exclusion of harms that have significant social and psychological consequences. A recurring question that emerged was whether intimate partner related harms should be considered a subset of cyberbullying or cyber extortion. The nuances of intimate partner harms, whether revenge porn or sextortion, make it difficult to categorize them in preexisting harm categories that make sense to consumers. Consumers may not be aware of differences between such harms, in recognizing that revenge porn involves an element of revenge and betrayal by an intimate partner, while sextortion may be perpetrated by anyone with access to sensitive data, through means such as hacking—and is dependent on ransom compliance. Both harms are exploitative in the same way, but the intent behind them varies. Similarly, although revenge porn would be more fitting as a subcategory under cyberbullying, there may be an extortionary element when coupled with expectations such as reconciliation-revealing how emotional coercion can function as a subtler form of extortion.

The growing prevalence of digital harms that impact individuals' mental health and social well-being calls for a recalibration of what is considered insurable. Including such harms in the scope of cyber insurance would not only enhance consumer trust but also provide better protection against the diverse array of risks consumers increasingly face.

4.2 Consumer Education

Given that people have a poor understanding of the actions associated with identity theft and cyber attack (49% and 53% correctly defined), the issue may not necessarily lie with the insurance policies but rather with a broader lack of consumer education. Providing consumers with better education about the nature of cyber harms could align their definitions with policy coverage, thereby increasing trust in the product.

Cyberbullying, appeared to be the most correctly defined harm, but this is complicated by the high rate of tautological definitions, such as "bullying online". While this phrasing suggests familiarity with the term, it lacks specificity and makes it difficult to assess whether participants genuinely understood the behaviors that constitute cyberbullying—such as repeated harassment or public shaming on digital platforms. It is unclear whether the prevalence of tautological definitions reflects a deep familiarity that renders further explanation unnecessary, or whether it indicates superficial understanding, shaped by the intuitive nature of the term itself. In either case, the frequency of tautological responses suggests that participants assumed a shared cultural understanding of what "cyberbullying" entails, making it hard to draw clear conclusions about their actual knowledge. The large share of tautological definitions may suggest the need for greater consumer education on what conduct cyberbullying entails and how it differs from harms such as cyber extortion—where threats can be seen as a part of cyberbullying, or a data breach-where disclosing or leaking data can be similarly be interpreted as cyberbullying.

Low understanding of cyber attack definitions suggests uncertainty surrounding the technical mechanisms (like system compromise or DoS) as being less intuitively grasped by the general public. Unlike more socially grounded harms like cyberbullying, which people may encounter or hear in everyday contexts, cyber attacks often involve specialized language and technical processes that are less visible and harder to narrate in lay terms. This points to a gap between public perception and the technical nature of cybersecurity threats. It also suggests that, despite its ubiquity in media and policy discussions, cyber attack definitions lack consistent and accessible meaning for many people. Strong understanding is critical

Our third research question provides insights into what that education should look like. Understanding of cyber perils appears to be distinct from education and income level, which is encouraging given improving both is a long-standing and difficult to achieve policy goal. It appears that improving security awareness will not improve understanding of perils from a cyber insurance perspective either, although this is based on correlational evidence. This suggests that targeted education campaigns are needed.

This education is likely best embedded within the insurance buying process. This could involve clearer communication of coverage, such as providing examples that illustrate the specific harms and potential scenarios covered by cyber insurance. It should also be considered whether participants should trust their own understanding of the risk landscape or defer to expert definitions which may differ from their lived experiences. This suggests an important role for insurance intermediaries like brokers to advise on personal cyber insurance coverage.

4.3 Limitations

Our study relied on participants providing definitions of cyber harms. Given individuals are under time pressure, the responses may not fully capture participants' real-world understanding of insurance coverages or their reasoning behind purchasing policies. This is especially true for outlining second-order impacts, which are not necessary for the underlying task of defining a peril. Future work could be designed to collect expectations of the costs associated with perils, as opposed to our design that focused on understanding of the peril itself.

Further, our analysis focuses on self-reported definitions rather than real-life claim scenarios. As a result, our results may not fully reflect how individuals interpret policy terms or whether participant defined harms would be paid out. Furthermore, despite achieving IRR through the coding process, there may still be some degree of subjectivity despite efforts to ensure reliability. While this paper includes participants from both the U.S. and U.K., cultural and regional differences may prevent generalizing from our findings.

5 Related Work

The topic of whether organizations who buy cyber insurance understand the coverage has not been studied in prior work [6–9], even though various insurance disputes over cyber losses have emerged [11, 20]. This question is even more important for personal cyber insurance,

given individuals lack legal expertise. However, only one article [14] has studied the demand side of personal cyber insurance.

Prior work has been focused on studying insurance products. Woods [3] studied the coverage and pricing of personal identity insurance, finding it covered a range of costs associated with response. Schutz et al. [1] explored personal cyber insurance coverage in Germany, finding it covers a range of first and third party coverages, as well as legal and IT services. Finally, Jain et al. [2] investigated cyber insurance coverage in the US and the UK under both personal cyber and home insurance policies. They found that home insurance covered digital devices damaged by conventional policies, but there was a lack of affirmative coverage for cyber losses, except from a minority of policies that covered identity theft. In contrast, the personal cyber insurance policies covered the six perils that were studied in this paper.

The study that captured the demand-side conducted a user survey exploring risk perceptions [14]. Participants were asked to estimate their own risk exposure. Participants believe that data breaches and cyberbullying happened with the highest frequency, meanwhile online frauds have the highest severity. They also found that the gap between participants and insurers' understanding of perils was "narrowest for identity theft" [14], which contradicts our findings. The prior study relied on participants' self-reported accuracy, whereas we directly verified it. This points to a fatal bias in researching user understandings via self-reported data—it is hard to detect when participants are confidently wrong.

6 Conclusion

We discovered discrepancies between the defining actions of cyber harms, and how they are defined by the general public. Participants are mistaken to varying degrees about what conduct individual harms entail, and thereby cyber insurance coverage for those harms. While cyberbullying was the most well understood harm, with the majority of participants correctly defining it, just under half of all participants correctly defined identity theft, most frequently conflating it with data breach. This outlines a societal problem that calls for increased education of online harms, and would lead to greater cyber insurance literacy. Identity theft coverage, that caused the most confusion, should be offered alongside data breach coverage in order to reduce the risk to consumers. There should be increased education efforts in the form of awareness campaigns and literacy programs that reduce the variation of actions identified for the most misunderstood harms. This would in turn translate to greater transparency and increased trust in insurance frameworks, and fewer claims denied, ultimately benefiting consumers.

Insurers should undertake educational initiatives to help consumers better understand the scope of coverage, particularly when it comes to less conventional or emerging harms. In this regard, Insurers may also consider revisiting their policy structures to incorporate new and evolving threats, such as revenge porn and nonconsensual imagery, which are currently overlooked, but the extent to which these changes would address consumer misunderstandings remains an area for future study. Shifting the conceptual framework of harms in policies and redefining what constitutes a "covered" harm would not only improve policy transparency but also ensure that policies provide adequate protection against a broader range of risks.

References

[1] Florian Schütz, Florian Rampold, Andre Kalisch, and Kristin Masuch. Consumer cyber insurance as risk transfer: a coverage analysis. *Procedia Computer Science*, 219:521–528, 2023.

[2] Rachiyta Jain, Temima Hrle, and Daniel W Woods. Insurance versus digital harm: a content analysis of home and cyber insurance policies in the usa and uk. *Journal of Cybersecurity*, 11(1):tyae031, 2025.

[3] Daniel W Woods. Personal identity insurance: Coverage and pricing in the us. *Journal of Financial Transformation*, 57:36–45, 2023.

[4] Sasha Romanosky, Lillian Ablon, Alexander Kuehn, and Thomas Jones. Content analysis of cyber insurance policies: How do carriers price cyber risk? *Journal of Cybersecurity*, 5(1), 2019.

[5] Daniel W Woods and Josephine Wolff. A history of cyber risk transfer. *Journal of Cybersecurity*, 11(1):tyae028, 2025.

[6] Martin Eling and Werner Schnell. What do we know about cyber risk and cyber insurance? 2016.

[7] Angelica Marotta, Fabio Martinelli, Stefano Nanni, Albina Orlando, and Artsiom Yautsiukhin. Cyber-insurance survey. *Computer Science Review*, 24:35–61, 2017.

[8] Gregory Falco, Martin Eling, Danielle Jablanski, Matthias Weber, Virginia Miller, Lawrence A Gordon, Shaun Shuxun Wang, Joan Schmit, Russell Thomas, Mauro Elvedi, et al. Cyber risk research impeded by disciplinary barriers. *Science*, 366(6469):1066–1069, 2019.

[9] Aggeliki Tsohou, Vasiliki Diamantopoulou, Stefanos Gritzalis, and Costas Lambrinoudakis. Cyber insurance: state of the art, trends and future directions. *International Journal of Information Security*, 22(3):737–748, 2023.

[10] Richard McGregor, Carmen Reaiche, Stephen Boyle, and Graciela Corral de Zubielqui. Cyberspace and personal cyber insurance: a systematic review. *Journal of Computer Information Systems*, 64(1):157–171, 2024.

[11] Josephine Wolff. *Cyberinsurance policy: Rethinking risk in an Age of ransomware, computer fraud, data breaches, and cyberattacks*. MIT Press, 2022.

[12] J David Cummins and Neil A Doherty. The economics of insurance intermediaries. *Journal of risk and insurance*, 73(3):359–396, 2006.

[13] George Loewenstein, Joelle Y Friedman, Barbara McGill, Sarah Ahmad, Suzanne Linck, Stacey Sinkula, John Beshears, James J Choi, Jonathan Kolstad, David Laibson, et al. Consumers' misunderstanding of health insurance. *Journal of Health Economics*, 32(5):850–862, 2013.

[14] Rachiyta Jain, Temima Hrle, Margherita Marinetti, Adam Jenkins, Rainer Böhme, and Daniel W Woods. " why would money protect me from cyber bullying?": A mixed-methods study of personal cyber insurance. In *2025 IEEE Symposium on Security and Privacy (SP)*, pages 1–27. IEEE Computer Society, 2025.

[15] Cori Faklaris, Laura A Dabbish, and Jason I Hong. A self-report measure of end-user security attitudes (SA-6). In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, pages 61–77, 2019.

[16] Jenny Tang, Eleanor Birrell, and Ada Lerner. Replication: How well do my results generalize now? the external validity of online privacy and security surveys. In *Eighteenth symposium on usable privacy and security (SOUPS 2022)*, pages 367–385, 2022.

[17] Sasha Romanosky. Examining the costs and causes of cyber incidents. *Journal of Cybersecurity*, 2(2):121–135, 2016.

[18] Lucy Qin, Vaughn Hamilton, Sharon Wang, Yigit Aydinalp, Marin Scarlett, and Elissa M Redmiles. " did they F*** ing consent to that?": Safer digital intimacy via proactive protection against Image-Based sexual abuse. In *33rd USENIX Security Symposium (USENIX Security 24)*, pages 55–72, 2024.

[19] Miranda Wei, Jaron Mink, Yael Eiger, Tadayoshi Kohno, Elissa M. Redmiles, and Franziska Roesner. Sok (or solk?): on the quantitative study of sociodemographic factors and computer security behaviors. 2024.

[20] Josephine Wolff. The role of insurers in shaping international cyber-security norms about cyber-war. *Contemporary Security Policy*, 45(1):141–170, 2024.

Demographic Category	% of Participants
Age	
18-29 years old	21.71%
30-39 years old	32.65%
40-49 years old	18.40%
50-59 years old	16.14%
60 or older	11.13%
Education	
High school	13.91%
Bachelor's degree	40.26%
Master's degree	15.40%
Doctorate	3.71%
Gender	
Female	54.54%
Male	43.81%
Household Income ($)	
Up to $25,000	5.38%
$25,000-$49,999	11.50%
$50,000-$74,999	9.46%
$75,000-$99,999	7.05%
$100,000 or more	16.33%
Household Income (£)	
Up to £20,000	7.61%
£20,000-£39,999	18.92%
£40,000-£60,000	11.69%
£60,000-£80,000	6.12%
£80,000 or more	5.94%
Employment Status	
Employed full-time	52.88%
Employed part-time	14.47%
Self-employed	9.65%

Table 3: Demographic breakdown of participants

A Codebook

The codebook was broken down into the following codes, each with a series of sub-codes.

A.1 Structure

Participant definitions were coded by structure, as either "explanatory" or "tautological". Definitions that elaborated on the harm defined were coded as "explanatory" as these provided specific examples or elaborated on specificities of the harm. Definitions such as "When someone steals your identity" for identity theft or "fraud online" for online fraud, were coded

as tautological, as they provide no insight on how the harm is perceived by the participant. Definitions were coded under one of the two qualifiers if participants did not define a harm or expressed uncertainty over its meaning.

A.2 Actors

For each definition, where applicable, we coded the victim and the adversary to identify patterns in how harms are perceived, focusing on the roles participants attributed to different types of victims and perpetrators. We distinguished between first-person, second-person, and third-person victims. Third-person victims were further categorized as individuals, (e.g., "an individual," "someone") or entities (e.g., "company," "organization"). Adversaries were coded under two third-person categories: generic – where an adversary is not clearly defined (e.g., "someone," "people," "individuals") or specific – indicating a perpetrator (e.g., "hackers," "cybercriminals," "unauthorized individuals").

A.3 Action

The action section of the codebook was developed by assigning each harm a subsidiary action that is most commonly associated with the crime. For example, identity theft was only assigned one code – impersonation, as there are no other actions directly associated with this harm. In contrast, cyber bullying was assigned multiple codes – public, private, harassment, discrimination, and "other". This takes different nuances of cyberbullying into account, such as whether it occurs in public spaces such as Twitter/X, or entails posting photos of someone on Facebook, as opposed to sending a victim messages privately.

Subcodes under each action differentiates subtle differences in angle through which participants view the conduct that causes the harm. For example, distinctions between stealing data and disclosing data are made under data breach, though personal data is compromised in both cases. Similarly, a distinction is made between cyber bullying that occurs once vs repeated, and discrimination where distinctions arise from social or identity based factors. This method ensures that the analysis acknowledges the complexity of harms, where the same type of harm can be perceived or experienced in varied ways depending on the circumstances, the nature of the attack, and its intended target.

A.4 Motivation

The motivation section of the codebook illustrates participants' perceptions of the end goal of each harm. For example, participants were more likely to provide motivation, specifically financial gain, when defining cyber extortion (48%), while fewer participants discussed motivations for data breach, with financial gain only being a motivation in 1% of definitions. Motivations for harms were divided into five categories, all of which were driven by malice: "financial gain," "impersonation," "blackmail/extortion," "information gain," and "destructive." These were coded when definitions included explicit references to the perpetrator's intent or purpose behind committing the harm. Any motivations falling outside these codes were coded as "other."

A.5 Technical Details

Definitions detailing technical details of a harm were coded under the 'technical details' category in our codebook, which consists of three subcategories. The first subcategory pertains to ransom threats, and includes four threat types that adversaries use to demand money from their victims: "lock," "encrypt," "publish," and "reputation." The second subcategory involves any social media platforms mentioned, predominantly used in cyberbullying definitions. The third subcategory identifies two main types of data addressed by participants: personal and organizational. Under the "personal" data type, there are two subcategories: "identifiers" and "financial information". The "identifiers" subcategory includes personally identifiable information such as names, email addresses, dates of birth, SSN, etc. The "financial information" subcategory covers credit card numbers and bank account details, among other information. On the other hand, the "organizational" data type has only one subcategory, "classified data", which refers to sensitive or restricted information held by organizations.

A.6 Impact on Victim

Finally, any mentions of the impact on a victim were coded. This included both direct impact, such as reputational damage or mental anguish, including response costs, such as legal fees or therapy costs incurred.