Multimedia Tools and Applications (2023) 82:12829–12858
https://doi.org/10.1007/s11042-022-14291-z

1224: NEW FRONTIERS IN MULTIMEDIA-BASED AND MULTIMODAL HCI

Reporting non-consensual pornography: clarity, efficiency and distress

Antonella De Angeli[1] · Mattia Falduti[1] · Maria Menendez-Blanco[1] · Sergio Tessaris[1]

Received: 12 January 2022 / Revised: 3 October 2022 / Accepted: 3 December 2022 / Published online: 17 January 2023

Abstract

According to recent legislative initiatives, non-consensual pornography is a crime in several countries and social media providers have a duty to provide their users easy to use mechanisms to report abuses. In this paper, we analyse the state of the art of the interfaces for reporting non-consensual pornography from the victim's perspective. Firstly, we analysed 45 content sharing platforms where aggressors might post non-consensual pornography. The analysis identified three distinct interaction styles for reporting the crime: *Scriptum* (a text-field where the user verbally describes the abuse), *Bonam* (a multilayered menu that includes a correct option), and *Malam* (a multilayered menu that does not include a correct option). Secondly, we conducted a within-subject study to evaluate the experience elicited by these interaction styles. Participants (N=39) were given a scenario and asked to report six blurred images as non-consensual pornography using a medium-fidelity prototype. The results exposed complex trade-offs between clarity, efficiency, and distress among the different interaction styles. These trade-offs open foundational research directions transcending boundaries between human-computer interaction and multimedia studies and interfacing computer science research with the law.

Keywords Image-based sexual abuse · Abusive content report · Revenge porn · Secondary victimisation

✉ Mattia Falduti
Mattia.Falduti@unibz.it

Antonella De Angeli
Antonella.DeAngeli@unibz.it

Maria Menendez-Blanco
Maria.MenendezBlanco@unibz.it

Sergio Tessaris
tessaris@inf.unibz.it

[1] Faculty of Computer Science, Free University of Bozen-Bolzano, Piazza Domenicani 3, Bolzano, 39100, Italy

1 Introduction

Non-consensual pornography is the hideous crime of sharing sexually graphic photos or videos without the consent of the individual portrayed in them [28]. The abuse is in constant growth and recently fuelled by the increased use of social media during the pandemic [61]. Non-consensual pornography has profound implications for the victims, their families, and the society at large. Yet, our knowledge of the phenomenon is still in its infancy and multidisciplinary research is needed to create legal [57], social [24], and technological safeguards [43]. This paper contributes to such a gap by studying (*i*) the user interface that mediates the notification of non-consensual pornography, and (*ii*) the experience that different interaction styles impose on the victims.

In Europe, where this research is situated, new legislation is introduced to regulate online behaviour, with a specific focus on hostile and criminal behaviours. For example, the EU Digital Services Act [26] requires all hosting service providers to ensure the availability of digital mechanisms for notifying illegal content. However, the design requirements for these interfaces are described at a high level of abstraction, in terms of accessibility and usability. Diversely, the legal requirements for reporting potential criminal activity are very specific. The EU Victims Rights Directive in recital 34 [27] places an emphasis on the clarity of the report while noting that *justice cannot be effectively achieved unless victims can properly explain the circumstances of the crime and provide their evidence in a manner understandable to the competent authorities*. Furthermore, this legislation signals that not everyone is equally at risk of being target of hostile behaviours, and emphasises the need to act against all forms of violence against women. Similarly, research shows that gender, race, and sexual orientation are dimensions that influence online harassment in social media [9].

Aggressors put victims in a vulnerable situation, which can be exacerbated when victims are required to describe the aggression to start a legal action. Reporting can be experienced as a secondary victimisation, which refers to an additional violation of the victim's legitimate rights or entitlements when they are requested to describe the aggression, mostly for legal purposes [54, 60]. The focus of our research is to investigate how interface design can ease the negative experience of reporting online non-consensual pornography from the victim's perspective. Pursuing this aim requires interdisciplinary expertise and skills; therefore, the research presented in this paper is the outcome of a collaboration between legal, data processing, and Human-Computer Interaction (HCI) experts. Combining legal and HCI research and methods, we analyse the usability of commercial interfaces, identify weak points and obstacles that stay in the way between the victim and an efficient report, and empirically investigate how different reporting styles affect user experience.

Two studies are presented. The first one describes an analytical evaluation of 45 commercial platforms that revealed both legal and usability issues. We identify three recurring interaction styles to report non-consensual pornography and refer to them as *Bonam*, *Malam* and *Scriptum*. Platforms under the *Bonam* interaction style allow users to accurately report non-consensual pornography by selecting a correct menu item. Platforms under the *Malam* interaction style presents a similar menu-based interface but in this case there is no item that accurately describes non-consensual pornography. Those under the *Scriptum* interaction style require users to describe the abuse in writing in a free-from text field.

The second study (N=39) compares and contrasts the three interaction styles in a user evaluation performed on a medium-fidelity prototype representing a content-sharing platform. The evaluation adopted the International Standard Organisation (ISO) [8] usability metrics, considering efficacy (completion and clarity), efficiency (time and clicks) and satisfaction (user experience). The results exposed complex trade-offs between clarity, efficiency, and distress. These trade-offs open foundational research directions for interface design, which transcend the boundaries between HCI and multimedia research as initially proposed in our contribution to CHItaly'21 [20].

The paper is organised as follows. Section 2 discusses multidisciplinary related work providing an introduction to the legal framework and technical research underlying the paper. Section 3 presents the analytical evaluation of commercial interfaces for reporting non-consensual pornography. The user study is presented in Section 4 including method and results. Section 5 summarises, compares and discusses results of the studies connecting them with related work. Section 6 concludes with a strong call for interdisciplinary research to improve the dramatic experience a victim of non-consensual pornography is likely to incur, when trying to exercise rights to dignity, sexual autonomy, and freedom of expression on-line.

2 Related work

Non-consensual pornography is the distribution of intimate, and sexual images or videos without the consent of the person depicted in them. Colloquially known as “revenge porn” (a highly misleading and potentially harming label that disregards the original context), non-consensual pornography is a form of image-based sexual abuse (see Fig. 1). Images can be real or fabricated. Real images can be stolen from the victim in physical settings or by hacking their devices. An example in physical settings is *upskirting*, which refers to the use of video cameras in public spaces to take intimate pictures without permission [51]. The act of hacking victim's devices is known as *Fappening*, a portmanteau of “The Happening” and “fap”, slang for masturbation [50]. It refers to the illegal release of an extensive online archive of hacked nude photos of - mostly women - celebrities. False images can be created with *Deepfake*, which refers to images or videos that portray unreal events as real through digital media manipulation [40]) or with *Deepnude*, which is the use of generative software based on image-to-image translation algorithms to undress photos of humans and produce realistic nudes [77].

The act of using and sharing sexual images without consent is categorised under different types of abuse. For example, *Sextortion* refers to the threat of distributing intimate sexual content unless a victim complies with specific demands [58]; *Cyberflashing*, is the act of sending unwanted sexual images or videos [30]; and *Non-consensual Pornography*, which

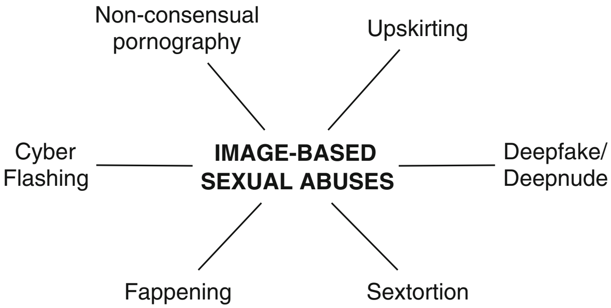


Fig. 1 Examples of image-based sexual abuse

usually refers to the use of images previously shared under a trusted relationship. The perpetrator is frequently an ex-partner who obtained the images during a relationship with the victim and now seeks retaliation for ending it. However, the motivation is not always to seek revenge, and the perpetrator can be a casual acquaintance, often met online. In any case, non-consensual pornography is committed with the intent to harm someone by shaming and humiliating them while publicly exposing one of the most intimate and private spheres of their life. Therefore, non-consensual pornography implies an invasion of privacy, removal of agency, and the violation of the human rights to dignity, sexual autonomy, and freedom of expression [65].

The current discussion concerning legal consequences and remedies is animated in several countries worldwide [14]. Italy, for example, has recently included the article n. 612-ter of Penal Code ("Illegal dissemination of sexually explicit images or videos"). It is formulated as follows:

> Unless the fact constitutes a more serious crime, a person who, after having made or stolen them, sends, delivers, transfers, publishes or disseminates images or videos containing sexually explicit materials, intended to remain private, without the consent of the persons depicted, is liable to be punished with imprisonment from one to six years and with a fine of between Euro 5,000 and Euro 15,000. The same penalty is applied to anyone who, having received or otherwise acquired the images or videos referred to in the first paragraph, sends, delivers, transfers, publishes or disseminates them without the consent of the persons depicted with the intention to harm them. The penalty is increased if the facts are committed by the spouse, even separated or divorced, or by a person who is or has been linked by an emotional relationship to the victim or if the facts are committed through IT or telematic tools.

Consequently, digital interfaces are key to enable clear and efficient measures to counteract illegal dissemination of sexually explicit content. The path to this crime (in legal terms, *iter criminis*) and the reaction after the crime is committed can be represented as a three-phases process: creation, distribution, and report (Fig. 2). In this paper we focus on the reporting phase and the user interface mediating it.

Non-consensual pornography brings forward multi-faceted issues and presents many multidisciplinary challenges. Psychological research has studied the perpetrator's motivation and how to support the victim [56]. Besides, the motivations for making and sending self-taken sexual images were investigated among different populations, such as young adults [15, 65] or people who self-identified as straight, gay, or lesbian [38]. Issues related

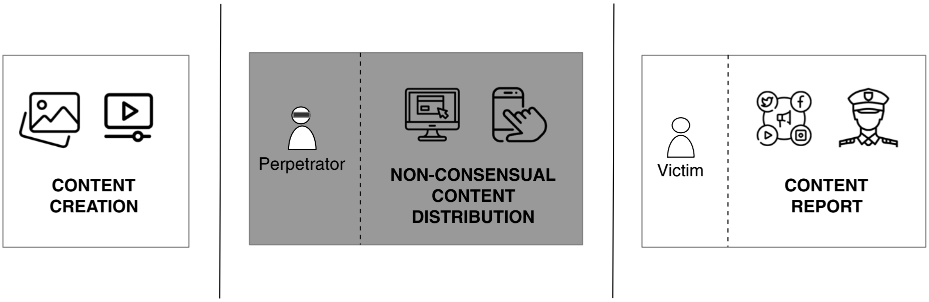


Fig. 2 The criminal flow of non-consensual pornography

to cyber-crime and in-person intimate partner violence are described in [49] and an analysis of multimedia sexting and computer-mediated sexual behaviour among young adults is reported in [24]. This paper expands this line of research by building bridges between computer science and the law because many legal issues must be taken into consideration when designing safe and respectful content-sharing platforms.

2.1 Legal implications

Despite most victims not reporting cyber-crimes to the police [74], non-consensual pornography has been under the spotlight for the toll it is taking on individuals and societies, with extreme cases linked to suicide. According to the Italian law enforcement agency [64], in 2021, there were more than 500 complaints with over 1.400 people accused. This represents an increase of almost 80% as compared to the previous year. Other European countries display more dramatic figures. We refrain in this paper from speculating on the reason for such difference. We simply notice that in Germany, the crime counted 9.233 cases in 2020 [13], and the UK Police registered 1,185 cases of illegally shared private multimedia content online in 2021. The press suggests that the pandemic has increased non-consensual pornography by 329% only in London [72].

The debate concerning non-consensual pornography has recently addressed the application of the principle of *notice and take-down* as regards platforms' liability [71]. Notice and take-down is a principle that allows private entities (i.e., individuals, companies, rights holders, and organisations) to interact formally with digital platforms in response to institutional allegations about a piece of media content that is illegal from a copyright law perspective. Accordingly, the platforms should remove the content or disable the access to it [1, 5]. However, institutional entities, both national and international (such as law enforcement agencies, courts, and public agencies), cannot monitor and moderate the enormous social media activities. This concern emerges clearly from a report of the North Yorkshire Police, which indicated that the law enforcement agencies were not providing support to the victims with practical issues, such as image removal [55].

Most service providers allow users reporting illegal or offensive content to address the task of regulating a vast and constantly changing collections of content. This system of self-regulation [66], often indicated as *flagging*, allows users to signal that a particular piece of media content is not appropriate and acceptable because it is against the "community guidelines" [17]. These guidelines are legally defined as terms of use [29] and serve as a contract between service providers and users. They are usually ignored by the average user [59], but even when read, they are of limited utility to the victim. Not only are they written in legal terminology, but they also require interpretation based on jurisdiction.

Digital platforms belong to providers who are real private entities, also existing offline. They have a nationality, a board, and ownership. Different nationalities imply different legal requirements, jurisdictions, and enforcement modalities. In some cases, even a seat in a country can influence the applicable law. Not all countries consider non-consensual pornography a crime, and the conduct included in the national legislation varies. Moreover, in some countries, similar illegal conducts are described differently, with different terms of penalties for the perpetrators and legal protection for the victims. Consistently, user notification is open to misunderstanding, disagreement, or misuse [42] but providers are under increasing pressure to "swiftly take down" illegal content once they are notified of it [35]. The EU Digital Services Act [26] elaborates this requirement in terms of usability and accessibility (Art. 14) by noting:

> [...] providers of hosting services shall put mechanisms in place to allow any individual or entity to notify them of the presence on their service of specific items of information that the individual or entity considers to be illegal content. Those mechanisms shall be easy to access, user-friendly, and allow for the submission of notices exclusively by electronic means.

This user-centred concern is echoed in the Online Safety Bill [35] under discussion in the UK. However, the provision of usable interfaces is not sufficient to ensure notification. Victims who have suffered humiliation and have been subjected to defamation of character might still resist reporting their experience. Fear of secondary victimisation can refrain victims from reporting non-consensual pornography [10]. In legal terms, secondary victimisation refers to the harm that occurs not as a direct result of the criminal act but through the response of institutions and individuals to the victim. This includes, but is not limited to, not recognising the psychological distress and treating them in a disrespectful way, an insensitive and unprofessional manner of approaching the victim, or a discrimination against the victim in any way [45].

Examples of gender harassment online and secondary victimisation from the digital platforms and the law enforcement agencies have been discussed concerning the denial of justice given to victims just after the crime commission [36]. The EU Victims Rights Directive [27], the bill that establishes minimum standards on the rights, support, and protection of victims of crime, affirms that

> (17) victims of gender-based violence and their children often require special support and protection because of the high risk of secondary and repeat victimisation, of intimidation and of retaliation connected with such violence. [...] (18) Member States shall ensure that measures are available to protect victims and their family members from secondary and repeat victimisation, from intimidation and from retaliation, including against the risk of emotional or psychological harm, and to protect the dignity of victims during questioning and when testifying. When necessary, such measures shall also include procedures established under national law for the physical protection of victims and their family members.

The Directive sets out goals that all Member States must achieve. It does not impose specific requirements on multi-jurisdictional platform owners. However, if we consider interfaces as “first responders” of a crime, the Directive directly talks to interface designers and social media developers. From this perspective, we can read Article 3 as a call for usable interfaces by recognising the victim’s *Right to understand and to be understood* and their *Right to receive information from the first contact with a competent authority*. Article 24 further elaborates on how to facilitate the process by recommending that

> When reporting a crime, victims should receive a written acknowledgement of their complaint [...] stating the basic elements of the crime, such as the type of crime, the time and place, and any damage or harm caused by the crime. This acknowledgement should include a file number and the time and place for reporting of the crime in order to serve as evidence that the crime has been reported, for example in relation to insurance claims. [...]

In this paper, we claim that interfaces may become unintentional partners in the secondary victimisation of people who have experienced non-consensual pornography and want to report it, and we call for synergies between the law and computer science research fields to avoid it.

2.2 Technical applications

The abundance of content daily shared on social media sadly implies cases of misappropriation or abuse [44]. Still, researchers working on technical solutions to online abuse have seldom addressed abusive behaviour from a user-perspective [6, 11]. Rather, research tried to counteract online abusive behaviours by developing technical approaches devoted to detecting problematic content [43] or improving digital forensics [76]. In current platforms, these ex-post actions are triggered by a user notification. However, the Web Foundation working group on online gender-based violence has identified several issues related to the mechanisms currently implemented for reporting and managing online safety. It is noted, for example, that i) victims are often not aware of the platforms policies and product features that can help them stay safe, ii) victims and social media platforms do not always describe or explain experiences of abuse as such, iii) online abuse can lead victims to stay silent and leave.

The relevance of considering users' needs in the design of practical solutions for preventing online abuses is highlighted in [67]. According to the authors, many studies concerning dissuasive strategies and technical solutions for mitigating non-consensual sharing do not place the user at the heart of the design process. They introduced a methodology for co-designing with end-users in the early stages of the design process and borrowed terminology and concepts from a crime-prevention legal framework used in the justice system to fill this gap. Moreover, a recent study on how dating apps could be designed to mediate the exchange of sexual consent and serve as a scalable solution to prevent sexual violence is presented in [33]. It describes an interface that encourages online discussion with potential sex partners about consent and sexual boundaries before meeting them face-to-face.

The legal framework revised so far provides several hints to interface designers. They encompass the entire space of the user experience, including usability, aesthetics, and symbolism [23, 37]. Yet, the task of reporting illegal content is hampered by two conceptual and practical challenges. The first one implies the definition of what is "illegal" content. Although it might seem relatively simple to reference illegal content as something contrary to the law, users must have access to the justice system to understand what is contrary to law. Nevertheless, besides being expensive and time-consuming, accessing justice is particularly difficult in the boundless space where social media platforms operate. What is illegal in some countries is not necessarily illegal in others.

The second challenge is related to communicating the occurrence of illegal content. As described in [57], removal of online images might be possible when the offender is known, and the victim is willing to file a complaint. Even under these conditions, there are emerging problems concerning i) the boundaries of the criminal relevance of various behaviours [14] and ii) the effectiveness of crime prevention strategies [52]. Preventive solutions face practical difficulties in automatically retrieving offending images [53]. Fundamental issues of affirmative consent on social platforms have been explored from a theoretical perspective and proposed as a possible solution for online harassment, non-consensual pornography, and problems with content feeds by [68]. The approach elaborated a theory combining feminist, legal, and HCI literature. Related research discussed strategies for implementing legal requirements in the design [34] and demonstrated that a victim-centred perspective could contribute to abuse prevention and mitigation [32].

To summarise, an emerging corpus of research has brought to the fore the need for expertise sharing between HCI and (criminal) law research [7, 20]. Most of this work focuses on holistic approaches to design. For example, [31] presents a qualitative analysis of the role technology design can play in intimate partner violence, while [70] describes how technologies can interact with justice-oriented service delivery in cases of violence against sex workers. Similarly, the opportunity for designs, policies, and algorithms to improve women's safety online is demonstrated by the results of a qualitative study of online abuse experiences [39]. However, minimal research investigated strategies for implementing legal protection in the design of interfaces. A notable exception is reported in [48]. The work describes a chatterbot for assisting and informing survivors of image-based sexual abuse. Using a chatterbot as the first responder to online crime is apparently well justified because chatterbots are "constantly available, do not judge the conversation partner, and may deliver structured information and words of comfort" [48, p. 1]. Preliminary results are encouraging, but previous research would suggest a note of caution as talking machines also tend to induce disinhibition and aggression in the user [12, 19, 22].

3 Analytical evaluation

Non-consensual pornography occurs on various multimedia platforms, such as social media, messaging applications, forums, image sharing websites, video sharing websites, blogs, live-streaming platforms, websites for sharing pornographic content and deepfake productions. Once the abuse is committed, victims can report the case to providers and law enforcement agencies. To better understand how victims can report these abuses on digital platforms, we conducted an expert-based evaluation of existing interfaces. The analysis addressed the following research question:

> *How complex is the process of reporting non-consensual pornography abuses from a victim-centred perspective?*

3.1 Sample selection

Despite a list of the most popular pornographic websites reported in [73], it is impossible to predict where perpetrators will commit crimes, considering the vastness of the web. Therefore, we started our study by identifying a sample of platforms to base the analysis. Specific attention was given to the context of South Tyrol, a province in the north of Italy in which this research is situated, where Italian and German are official languages. We engaged in the following four activities:

1. We identified an initial list of platforms where non-consensual pornography had occurred based on the analysis of popular cases reported in the news and the scientific literature (e.g., [41] and [14]);
2. Focusing on those platforms, we followed cross-references in posts, news, and blogs where the topic of non-consensual pornography was discussed, adding the names of new platforms to the list;
3. We iteratively expanded the list with similar platforms from the point of view of the user base or primary service;
4. We included the most popular video and image sharing websites for adults identified using the following search strings on Google *most popular porn websites* and *most popular deepfake porn websites* in English.

Data collection was interrupted when reaching saturation meaning that cross-references did not add additional relevant information or the links pointed to landing pages with ambiguous advertisements. Through this process we identified 45 platforms, which we classified according to their primary service based on content (adult/explicit or family-friendly) and media hosted (text, images, and/or videos). For adult/explicit material we intended media content representing erotic behaviour or where intimate body parts representing erotic behaviour were not censored. Access to these platforms always required confirmation of age.

The same platform appealed to more than one audience or provided more than one service in several cases. In these cases, coding reflected the primary goal advertised on the platform website. For instance, platforms were coded as social media if they allowed users to share text, images, and videos as their core business (e.g., Facebook). However, when the primary service was connected to a specific media (such as video sharing for YouTube), we classified the platform accordingly. At the end, the sample was composed of 25 family-friendly platforms and 20 platforms dedicated to adult content. Family-friendly platforms included messaging apps (8), social media (5), image sharing and image boards (3), video sharing and forum (2), live-streaming sharing and micro-blogging (1). Platforms for adults included porn video sharing (12), deep fake porn sharing (4), porn forum (3) and deepfake forum (1). Considering the illegal nature of several websites, we omit to disclose the complete list in this paper, but we are willing to share it upon request for research purposes.

3.2 Results

A legal expert specialised in criminal law and cybercrime and with a PhD in computer science analysed all the platforms taking into account their *legal characteristics* (terms of use and jurisdiction) and *interaction design* (availability of reporting interfaces, precision, and style). The legal analysis was performed individually, with double coding only when the information could not be found. Interaction design variables were analysed by two researchers who performed a systematic walk-through of all the platforms to simulate a report of non-consensual pornography. They recorded all steps of the procedure and their success. Tasks were scored as successful when the researchers managed to file a complaint, and the submit button appeared. For obvious ethical reasons, the task was interrupted at this point. Hence, we have no evidence of effective delivery or follow up services. Tasks were scored as unsuccessful if the researchers could not perform the task after extensive testing. Once the optimal procedure was identified, we counted the minimal number of actions required for optimal performance and clustered the platforms according to interaction style and report precision.

3.2.1 Legal characteristics

Each platform's legal characteristics were inspected considering its terms of use and jurisdiction. All but two of them provided terms of use page. Around 60% allowed sharing pornographic content, and only one of them explicitly forbade illegal pornography involving children or private media. The rest of the platforms (40%) explicitly forbade posting, uploading, and sharing any form of adult content. The terms of use were analysed by considering the document's language, concerning the language spoken in the context where the study was performed (German and Italian) and English as the most common language in digital contexts.

The analysis showed that 25 platforms provided multilingual terms of use, two platforms provided bilingual terms of use (English and Italian), and 16 platforms provided monolingual terms of use in either English (10) or Italian (6). The analysis suggested that the presence of multilingual documents was influenced by the providers' market position, size, seat, and user base. Indeed, small and nationally-relevant providers did not support a multilingual service, whereas multinational big tech companies provided access to several translations.

The provider's jurisdiction was inferred by the contact details listed on the platform. Such formal details are essential for the victims because they may need to contact, interact, and establish a connection formally with the providers. In addition, they frame the legal space where the platforms operate and the law they must obey. However, around 36% of the providers did not disclose any contact details. The rest were distributed across the northern hemisphere: the USA (17), Canada (5), Japan (3), Italy (2), Czech Republic (2), China (2), and Russia (1).

3.2.2 Interaction design

The researchers were unsuccessful in reporting the content to the provider of some 36% of the platforms. In only 16% of the reports, they could precisely describe the occurrence of non-consensual pornography. When available, reporting interfaces were based on a menu-based interaction (N=17), where the user could select items to denote the event, or a free-text form (N=8), where the user could describe the event with their own words. Menu-based interfaces differed according to the precision of the items. They could make an unambiguous reference to the non-consensual sharing of intimate images or provide more general categories such as nudity, pornography, or harassment. Consequently, we identified three main interaction styles that differed in precision and interface widgets. We named these styles as *Bonam, Malam* and *Scriptum*. The label choice for *Bonam* and *Malam* is based on "in malam partem" and "in bonam partem", respectively, both Latin expressions commonly used in the criminal law domain. *Scriptum* was selected as the literal translation of "piece of text" in Latin and declined in the accusative singular for consistency with *Bonam*, and *Malam*.

- **Bonam** interfaces are based on a nested menu, which includes an unambiguous item to report intimate images shared without content. The example available on Facebook is `Submit a Report > Involuntary pornography`.
- **Malam** interfaces are based on a nested menu, which provides a general label to indicate the abuse. The example available on TikTok is `Report > Pornography and nudity`.
- **Scriptum** interfaces require the user to provide a written narrative of the abuse in an free text field, which could be accessed through a nested menu. The example available on Youtube is `Report > Sexual Content > Nudity > Provide additional information`.

We counted the number of clicks necessary to submit the report for each interaction style and computed their average. As we can see in (Table 1), on average *Bonam* requires an extra click to be completed as compared to *Malam* and *Scriptum*. Table 1 presents the variability of interactions necessary for notifying non-consensual pornography. Differences addressed the terminology used, the number of interfaces the user has to navigate (range 2-6) and report's style (selection or narrative). It is important to notice that these interaction

Table 1 Design characteristics of interfaces for reporting non-consensual pornography

Interaction Path	Length	Type	Average
Report abuse + Intimate content posted without my consent	2		
Menu + Report + Non-Consensual Sexualisation	3		
Report + Other issues + It's involuntary pornography + I appear in the image	4		
Menu + Find support or report photo + Nudity + Sharing private images	5	Bonam	4.3 (SD=1.6)
Menu + Report + It's inappropriate + Nudity or sexual activity + Sharing private images	6		
Menu + Embed + Report abuse + Privacy violation + Yes, my privacy + My explicit images	6		
Report + Infringes My Rights	2		
Report + Pornography/Nudity	2		
Menu + Report + Offensive/Copyright/Other	3		
Menu + Report post + Personal information or fake photo manipulation	3		
Menu + other + Porn content	3		
Menu + Report + Report Abuse	3	Malam	3.3 (SD=1)
Menu + Report + Sexual Harassment	3		
Menu + Report + Nudity + Regarding Me	4		
Menu + Report Tweet + It's abusive or harmful + Includes private information + hacked materials	5		
Menu + Report + Sexual Content + Nudity + Select Timestamp	5		
Report + Free text	2		
Report + Choose (1x) + Free text	3	Scriptum	3.5 (SD=1,7)
Menu + Report + Free text	3		
Menu + Report + Choose Option (3x) + Free text	6		

styles do not only differ from a usability perspective but, more importantly for the purpose of this paper, they are crucially different in terms of their legal implications. For example, if a user flags an intimate sexual photo shared without consent as "pornography and nudity", which might be the closest option in the *Malam* interaction style, then the report signals that the media content is sexually explicit or representing exposed genitals, but legally legitimate. However, if a user flags a media content with the "non-consensual pornography" label available in the *Bonam* option, the legal implications are drastically different. In this case, the report signals that the media content is not only sexually explicit but also abusive and intended to harm the victim. These connotations are crucially different from a legal perspective because they constitute an abuse and can therefore be reported and consequently prosecuted.

4 User study

A medium-fidelity prototype was designed to compare the experience elicited by the three interaction styles for reporting media content. This level of fidelity brought forward two key benefits to the study. First, it allowed the participant to focus on the pragmatic task of reporting the abuse rather than on the aesthetics and the symbolism of the interface. Consequently, it avoided that attitudes towards the familiarity with a brand influenced the evaluation [21]. Second, it allowed for evaluating performance and judgement without exposing the user to any sensitive content, as clearly come out in our discussions about the ethical considerations of our research. These discussions were informal, as a personal stance towards research, and formal since we needed to prepare an application for the Ethics Review Board of the Free University of Bozen-Bolzano, which approved the study.

Aligned with contemporary approaches to research ethics in user studies, we considered several situations that could make the participants feel uncomfortable. For example, in addition to a written consent form, one of us explained the study verbally to each participant, making sure they felt comfortable with the topic and would quit at any time if they wished to do so. Besides, we carefully designed the task not to prime the gender of the victim and used blurred images to evoke a body without revealing any specific part or colour.

Participants were recruited through convenience sampling [25] targeting university students and younger adults who represent the typical social media user [3] and are more likely to have experienced cyberbullying in their life [44]. However, as the number of older users is rapidly increasing [16], no age requirement was imposed on participants above 18. The target group was selected based on their knowledge of English since this was the language in which was the study was conducted. All the students in the study were engaged in English-based education and the other participants were selected based on their English skills.

4.1 Method

4.1.1 Participants

A total of 43 participants accessed the experimental prototype, but four did not perform any task. Consequently, data were available for 39 participants (21 men, 16 women, one non-binary, and one not declared) and 234 tasks. The majority of the participants (51%) reported to be aged between 25 and 34 years old and 23% between 18 and 24. The remaining sample was evenly divided in the following decades, with three participants over 65 years old. Around 42% of them had real-life experience with reporting a large variety of online content, including nudity, underage pornography, harassment, spam, scam and frauds, ethnic hate and racism, or bullying and violence. One-third of the reports addressed images as abusive content.

4.1.2 Design

The study applied a within-subjects design with Interaction Style (3) as repeated measures. Participants were exposed to each interaction style (i.e., *Bonam*, *Malam* and *Scriptum*) two times and performed six reporting tasks using the medium-fidelity prototype.

- The *Bonam* style presented a nested menu, organised in two levels. Level one included ten items, and level two had 47 items. There was one correct option, namely the one presented in Facebook `Something Else > Non-consensual intimate images`.
- The *Malam* style presented a nested menu, organised in two levels. Level one included 14 items, and level two had 19 items. There was no correct option. Instead, the content

could only be generally flagged as pornography or nudity. The lower number of items, compared to *Bonam* was chosen not to overload people who looked for the correct label.

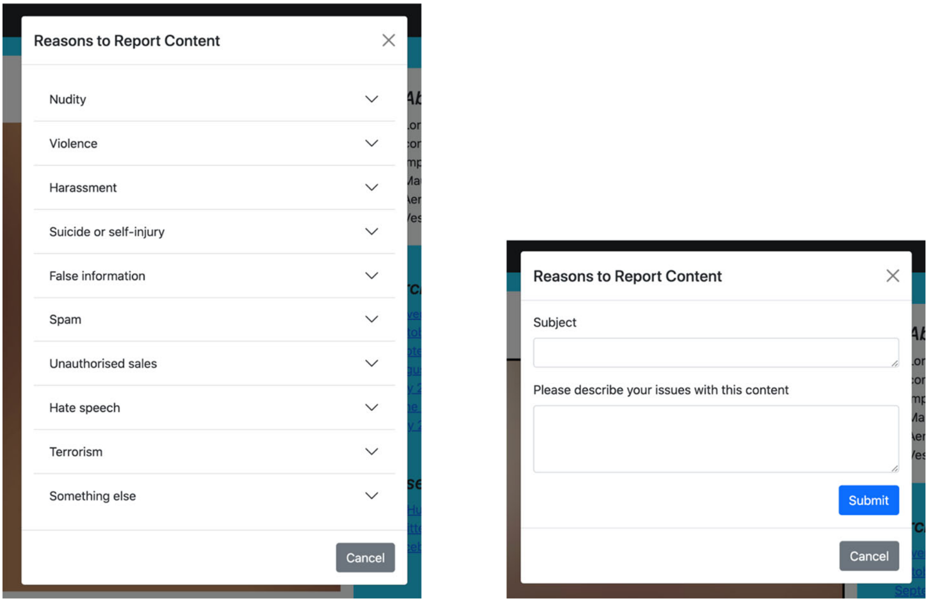
- The *Scriptum* style required the participants to produce a free-text narrative describing the abuse. The prototype did not require the user to enter any personal data (e.g., email address) or activate an account, as most digital platforms do.

The structure of the menu displayed in *Bonam* and *Malam* is available in the appendix. It was inspired by the real examples implemented in commercial platforms and identified in the analytical evaluation. The interface appearance for Bonam and Scriptum are depicted in Fig. 3. Malam appeared exactly like Bonam, but it contained different items.

4.1.3 Medium-fidelity prototype

The prototype resembled a generic blog (Fig. 4) in such a way that the interface was familiar without recalling any specific provider. The style followed commercial standards regarding menus, icons, and textual information. A red flag icon (in the right corner of the image window) activated the reporting process on all interaction styles. The difference between them was the reporting window, which could be based on a menu (for *Bonam* and *Malam*) or free text. The prototype was implemented using standard web technologies.

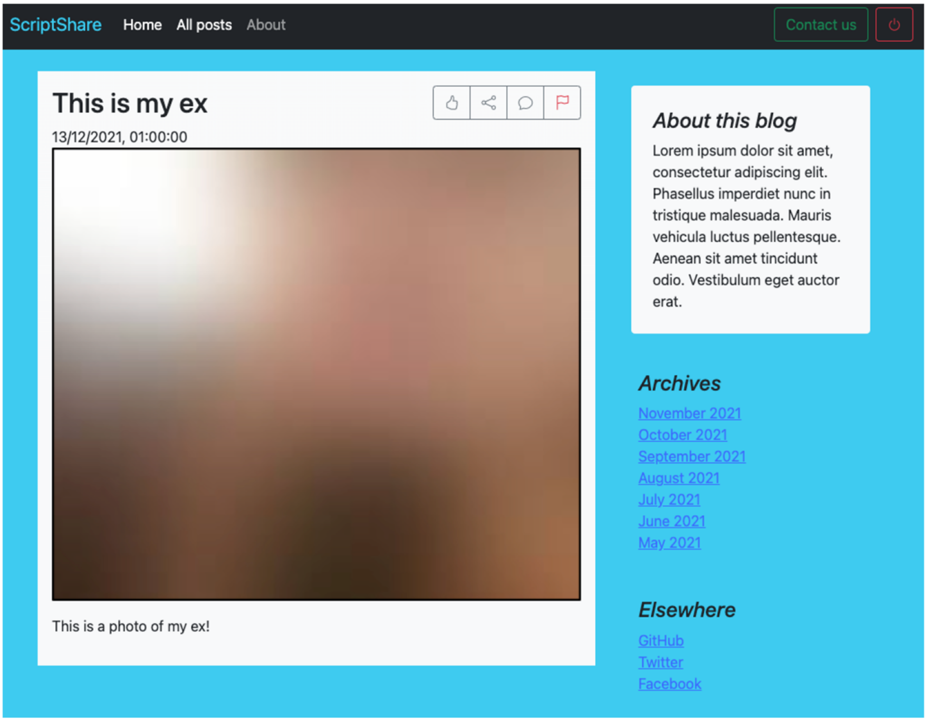
The software was composed of two modules: the controller that orchestrated the experiment flow and collected the data and the interfaces that participants were asked to evaluate. Both modules were deployed as web applications served by a dedicated HTTP server; however, while the controller relied on a Python Web Server Gateway Interface (WSGI)



(a) Bonam

(b) Scriptum

Fig. 3 Prototype reporting windows



This is my ex

13/12/2021, 01:00:00

This is a photo of my ex!

About this blog

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus imperdiet nunc in tristique malesuada. Mauris vehicula luctus pellentesque. Aenean sit amet tincidunt odio. Vestibulum eget auctor erat.

Archives

November 2021
October 2021
September 2021
August 2021
July 2021
June 2021
May 2021

Elsewhere

GitHub
Twitter
Facebook

Fig. 4 Layout of the medium-fidelity interface

framework[1], the interface prototypes were served as static sites with a Jamstack-like architecture.[2] To maximise the participant privacy, data were collected only on the interface prototypes using JavaScript embedded in the web source and an API running on an institutional web server. Scores were saved only after the completion of the experiment. If participants decided to interrupt the study, the data were not sent to the server. Consequently we avoided the deployment of the prototype on a centralised web server. Both *controller* and *the interface* were served by a simple HTTP server that each participant needed to install on their local computer. The source code is available at [4].

4.1.4 Procedure

First, we conducted a pilot study with an expert in experimental user research who is not related to the project. The outcomes helped improving a few aspects of the study such as the readability of the project description and the screen size. Then, we conducted the user study, which took place in the university rooms, and in private houses. First, one of the researchers welcomed the participants, explained the general goal of the study and invited them to ask any questions or provide comments to create a comfortable atmosphere. Then, the researcher shared the link to the online study, which the participants accessed through their own laptops or desktop computers. A schematic representation of the online study is

[1] https://www.python.org/dev/peps/pep-3333/

[2] https://jamstack.org/what-is-jamstack/

depicted in Fig. 5. Participants reached the consent page, which included a brief description of the research aims and the procedure. They were informed that the focus was the system's performance and not their capability and that they could interrupt the study at any point without consequences. Besides, they were ensured that all data were stored anonymously, and an e-mail address was provided for further information.

Upon agreeing to participate, the participant was asked two questions. The first one addressed whether they had ever reported offensive content via a multiple choice question (i.e., I prefer not to say, I don't remember, Yes, No). If the participant answered yes, they were invited to elaborate on the experience in an open-ended field. Then, the participants were presented with the following scenario, inspired by real cases reported in the news [41] and the literature [14].

> Imagine to be Alex. Alex is your age and is browsing through their social media. Unexpectedly, an intimate image appears on the homepage with the following title: "This is my ex". Alex recognises the private situation in which the picture was taken, and realises that the image has been uploaded without consent. Alex wants to report the abuse to the platform immediately. Your task is to **report the post** using the following interfaces.

Six tasks (two for each interaction style) were presented to each participant in random order. The participants had to report a blurred image, and the reporting modalities randomly differed according to the experimental condition (Fig. 4). The randomisation was automatically implemented in the prototype. At the end of each task, a questionnaire was presented. It consisted of seven semantic differential items and an open-text question where participants could elaborate on their experience. The closing page proposed three additional questions about demographic data (i.e., age range, gender, and academic background) and thanked the participants while providing a contact e-mail.

4.1.5 Measures

Following the ISO [8] the study took into consideration three usability dimensions: efficacy, efficiency, and satisfaction. Efficacy was defined as task completion and report clarity. The former referred to the number of participants who submitted the notification. The latter was evaluated by the quality of the report of being coherent, intelligible, clear, and easy to understand. Efficiency scores included the time and the number of clicks from flag selection to report submission. Consistently, it was computed only for the participants who completed all the tasks (N=33). Scores on the two tasks were averaged for each participant, condition and task. Finally, satisfaction was measured by an ad-hoc questionnaire designed

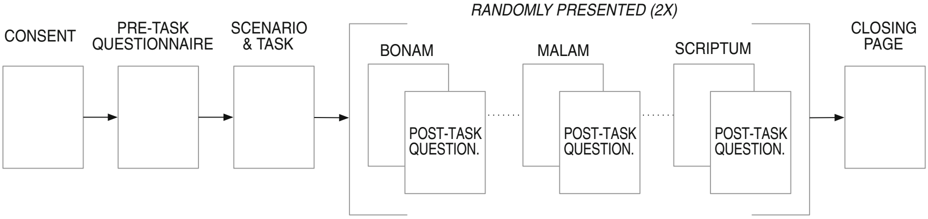


Fig. 5 Schematic representation of the user study's pages and tasks

for the study. It included both pragmatic and emotional judgements in seven semantic-differential items: easy-difficult, embarrassing-comfortable, quick-slow, clear-confusing, safe-dangerous, demanding-simple, and unsatisfactory-satisfactory.

4.1.6 Thematic analysis

As discussed in the related work section, defining non-consensual pornography at a multi-national, multi-language and multi-jurisdictional level is complex and sometimes even controversial. However, a definition of non-consensual pornography is available in the literature and several legal systems as the distribution of intimate images without the consent of the person represented in it. Based on this definition, two of the authors reviewed each report putting themselves in the position of a person who received the notification and needed to understand the nature of the problem to take legal action. After this initial analysis, we developed three categories to assess the level of clarity of the report:

- *Full*: the report contains both a mention of intimate images and lack of consent.
- *Partial*: the report contains either a mention of intimate images or lack of consent.
- *Poor*: the report contains no mention of intimate images or lack of consent.

Then, both researchers analysed individually all the reports following a deductive thematic analysis [69] and coded them into one of the three categories. The outcome of the analysis was discussed among the co-authors and disagreements were settled to reach consensus. Then, two of the authors deductively coded the participants' replies to the open-ended questions into the three usability dimensions, namely efficacy, efficiency, and satisfaction.

4.1.7 Results

Clarity For the tasks conducted under the *Bonam* interaction style, reports were labelled as full if the correct choice was selected, meaning `something else > non consensual intimate images`. A total of 11 reports were labelled as full in the *Bonam* condition. Tasks conducted under the *Malam* condition could not be label as full since, by definition, no specific item was available to refer to non-consensual pornography.

For the tasks conducted under the *Bonam* and *Malam* interaction styles, reports were labelled as partial if the selected option contained some key aspects but not all. For example, `nudity > sharing private image` was an option that our participants selected often but we labelled as a partial report since there is a reference to the intimate aspect of the content but not to the lack of consent. A total of 43 reports were labelled as partial in the *Bonam* condition, and 40 in the *Malam* condition.

Finally, reports labelled as poor did not refer to any of the intimate or non-consensual aspects of the content, as when the participants selected the option `harassment > me`. A total of 23 reports were labelled as poor in the *Bonam* condition, and 36 in the *Malam* condition. The list of items selected in the two menu-based conditions is reported in Table 2.

For the tasks conducted under *Scriptum*, we analysed the textual production and labelled it depending on the extent to which the text reported the intimate and non-consensual aspects of the content. Examples are *This picture shows an intimate situation and was uploaded without consent* (full); *I found an image uploaded without the consent of the person depicted* (partial); and *I'm in this picture and I don't like it* (poor). In this way, a total of 22 reports were labelled as full, 38 as partial, and 17 as poor in the *Scriptum* condition. The distribution of the reports as a function of clarity and interaction style is illustrated in Fig. 6.

Table 2 Options selected in Bonam and Malam

Bonam	**Frequency**	**Clarity**
Something else, non-consensual intimate images	11	Full
Nudity, sharing private images	33	Partial
Nudity, adult nudity	6	Partial
Something else, sexual exploitation	2	Partial
Nudity, sexual activity	1	Partial
Nudity, sexually suggestive	1	Partial
Harassment, me	6	Poor
Harassment, a friend	4	Poor
Violence, graphic violence	3	Poor
False information, something else	2	Poor
Something else, mocking victims	2	Poor
Something else, sharing private images	2	Poor
Hate speech, sex or gender indentity	1	Poor
Nudity, involves a child	1	Poor
Spam	1	Poor
Violence, something else	1	Poor
Total	77/78	
Malam	**Frequency**	**Clarity**
Pornography and nudity	40	Partial
Harassment or bullying, me	10	Poor
Harassment or bullying someone I know	7	Poor
Other	5	Poor
Harassment or bullying, other	4	Poor
Violent and graphic content	4	Poor
Misleading information, other	3	Poor
Illegal activities and regulated goods, promotion of criminal activities	1	Poor
Spam	2	Poor
Total	75/78	

Efficacy The large majority of the tasks (97%) resulted in the successful submission of a report. However, six participants (15%) did not complete all the tasks. Failures mainly occurred during the first (N=5) or the second task (N=2). Five participants did not select the flag icon, which activated the reporting procedure and therefore did not discover how to

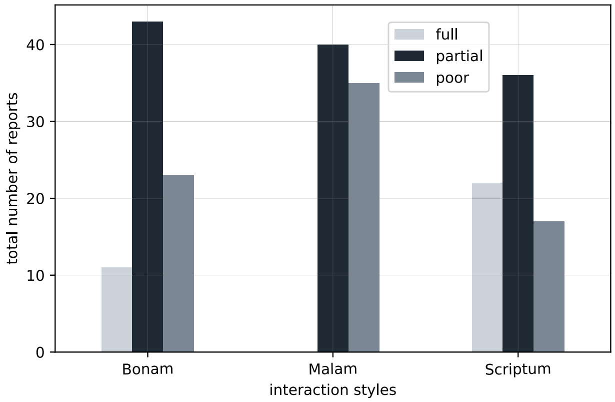


Fig. 6 Clarity distribution as a function of Interaction Style

execute the task. Instead, the other one correctly filled in the report but failed to submit it. With repetition, however, all the participants were capable of reporting offending images. Independent of the condition, the majority (89%) selected the flag with just one click. Before finding the flag, the other participants tried different links, from 2 (N=2) to 9 (N=1).

Non-parametric tests were computed separately on the scores of the first and the second task to compare *Bonam* and *Scriptum* where *full* clarity was possible. A Wilcoxon Signed Ranks test showed a significant effect for the first task $Z = (-1.98)$, $p < .05$, but not the second. Initially, participants using *Scriptum* were more accurate in reporting the abuse. In the second task, instead, the effect disappeared due to a slight improvement in the number of full reports in the *Bonam* condition (+ 8%). In the *Malam* condition, where full clarity was impossible by design, almost half of the reports were considered poor because they did not include any of the critical elements of the abuse (i.e., pornography and lack of consent).

Efficiency Efficiency was computed for all completed tasks, independent of their clarity (Fig. 7). It considered the time spent, and the number of clicks from flag click, to report submission. The mean duration of the study was 17 minutes (SD=9). For each task, time scores were averaged, normalised with a logarithmic transformation, and entered as dependent variables in a repeated measure ANOVA with Interaction Style as the within-subjects factor. A significant effect emerged $F_{(2,64)} = 18.52$, $p < .001$. The analysis of the contrasts demonstrated that *Scriptum* was the longest method, followed by *Bonam* and *Malam*, which were not significantly different from each other.

The mean number of clicks was 8.2 (SD=4.24). Figure 8 illustrates the average number of clicks performed by the users for finding the flag in the first task and submitting the report in all tasks as a function of interaction styles. For reference, we visualise these values alongside a measure of the best performance (the minimum number of clicks necessary to complete the task). The analysis of the number of clicks demonstrated a more nuanced frame of results. The variable was normalised for each condition, task and participant, considering the minimum number of clicks necessary for optimal performance. In this way, the best performance was represented by a score of 1 and negative scores increased as a function of errors. The values of both tasks were averaged and entered in a Friedman two-way analysis

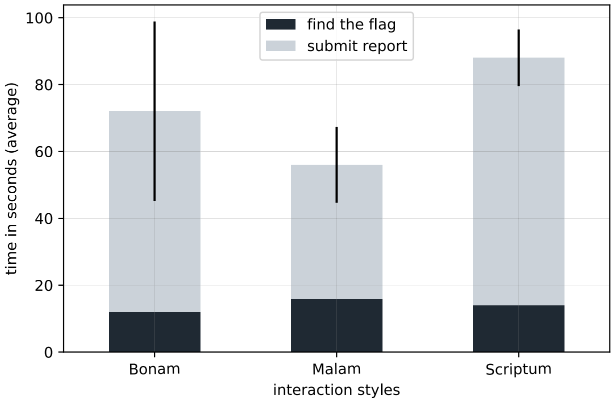


Fig. 7 Performance time as a function of interaction style

of variance with Interaction Style as the within-subjects factor because the data distribution did not satisfy normality requirements. Results highlighted a significant effect of Interaction Style on the number of clicks $\mathcal{X}^2 = 22.34$, $p < .001$. *Scriptum* required the lowest number of clicks, followed by *Bonam*, and *Malam*. A Wilcoxon Signed Ranks test was used to compare the number of clicks in the two menu-based conditions demonstrating a significant difference in favour of *Bonam*.

Satisfaction The results of the post-task questionnaire provided quantitative and qualitative data to understand the user experience. Reliability and factor analysis proved the unidimensional nature of the semantic differential scale in each condition, with all Cronbach alpha values $> .88$ (see Table 3).

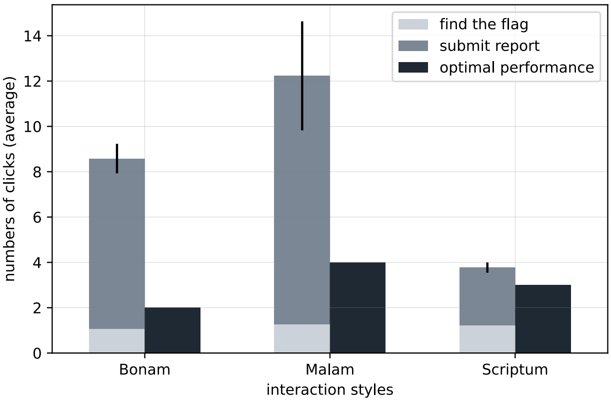


Fig. 8 Average click as a function of interaction style

Table 3 Corrected Item-Total correlations for the UX scale in the interface condition (Cronbach alpha > .88). R=reverse coded

Item	Bonam	Malam	Scriptum
ScriptumEasiness	.755	.689	.841
ScriptumEmbarrassment (R)	.573	.553	.743
ScriptumQuickness	.720	.755	.680
ScriptumClearness	.686	.753	.608
ScriptumSafeness	.482	.585	.577
ScriptumDemanding (R)	.750	.730	.807
ScriptumUnsatisfaction (R)	.662	.723	.686

Consistently, a single UX-index was computed by averaging scores of the seven items and then entered as a dependent variable in a repeated measure ANOVA with Interaction Style as the within-subjects factor. Results returned a significant effect for Interaction Style $F_{(2,76)} = 16.81, p > .001$. Pairwise comparisons indicated that the effect was due to the negative evaluation of *Scriptum*, whereas no significant differences emerged between the two menu-based conditions. Descriptive statistics are reported in Fig. 9.

The thematic analysis of the replies to the open questions allowed us to dig into the user experience. Distress was a major theme in the *Scriptum* condition. Many participants mentioned that *Scriptum* caused embarrassment while requiring the user to describe the abuse. Negative emotions were elicited by the sensitive nature of the topic and the awareness of a remote interlocutor. Example citations are reported below.

> It was embarrassing having to describe the subject and the issue. I felt negatively inclined to write text, as it felt I would need to write a lot and even worse when I started describing the image due to the nature of the content. [P12, Scriptum]

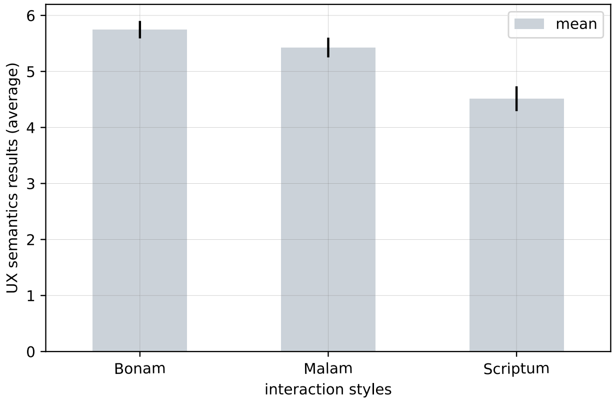


Fig. 9 UX-index as a function of Interaction Style

> This kind of report really makes me uncomfortable describing my feelings about the picture I am reporting. [P11, Scriptum]

> Its embarrassing to think that someone will read that. [P3, Scriptum]

In addition, participants were concerned about the effectiveness of their own report in *Scriptum*. They appreciated that it was possible to describe the issue in their own words, but they also found it difficult to find the right words to describe the abuse, which made them wonder whether their description would be efficient.

> It was not so easy to find the right words to describe my issue. I hope the image will removed soon. Impersonal and I was left with the feeling that the picture could not be removed / my report could be useless. [P7, Scriptum]

Conversely, participants appreciated the predefined categories available in *Malam* and *Bonam*. They found the menu-based conditions "fast and intuitive" [P25, Malam]. Besides, it increased perceived efficiency.

> I got the feeling I could describe or at least label my issue, I have the hope that this fact can make the process of removing my picture quicker and more urgent. [P7, Bonam]

On the negative side, some participants complained that scanning through the long list of options available in *Malam* and *Bonam* was not only time-consuming but also confusing. Using the *Bonam* interface, these drawbacks were partially compensated by the existence of an option that seemed accurate.

> Definitely optimal. Super fast and easy, once a bit acquainted with the interface also super immediate. All the information required is relevant, necessary, clear and the system doesn't expect the user to spend more than necessary reporting an uncomfortable item. [P2, Bonam]

> Just felt good to report the post clearly like this. It felt a good report even if was quick and simple. The report was specific but simple to read. I like this one. [P42, Bonam]

Conversely, some participants in the *Malam* condition expressed a feeling of confusion and dissatisfaction. The fact that they could not find an option that entirely matched the content they wanted to report made them feel helpless and left them thinking that they had not been able to describe the abuse accurately:

> The experience was ok. Highly unsatisfactory due to the generality of the options. A lot of ambiguity, and some categories inferred to the same concept. "Other" section utterly useless. Definitely fast enough, doesn't expose the user reporting and keeps the task simple quick and comfortable. [P2, Malam]

> Didn't find the right one, just reported in a general tab. Not sure which one was the right one. The report list was good but it missed a tab that seems the right. [P42, Malam]

> This interface does not let me describe the reason why I am reporting the content. [P28, Malam]

> I put "other" because I couldn't find the option I was looking for (private content without my consent). [P35, Malam]

5 Discussion

The paper reported two studies that investigated the state of the art concerning commercial user interfaces for reporting non-consensual pornography. Results demonstrated a very complex situation that calls for interdisciplinary research including criminal law and computer science. The first study identified and systematically analysed 45 platforms, representing possible venues where the crime may occur. They included a variety of technology and media targeted to family-friendly or adult content. Almost 4 out of 10 platforms did not have any reporting interfaces and only 16% allowed users to indicate the occurrence of non-consensual pornography with a proper legal vocabulary. Three key interaction styles were identified based on interface widgets and the precision of the items contained in them. The first style *Bonam* provides a menu-based interface containing a specific option to denote non-consensual pornography. The second, *Malam*, consists of a menu-based interface not containing a specific option to denote non-consensual pornography. The last one, *Scriptum*, is a free-text widget where the user can provide a written narrative.

The second study compared the usability of these interaction styles in a user evaluation performed on a medium-fidelity prototype. Results demonstrated multifaceted complexities in reporting non-consensual pornography online, starting from the identification of the reporting methodology and ending in the articulation of clear and unambiguous report. Despite the prototype following established standards in blog design, some 15% of the participants could not report the abuse the first time they met it. A common problem was related to the selection of the *flag* which activated the digital procedure. The problem may be due to the icon salience and its meaning. Participants may have failed to *see* the flag, which, in the prototype, did not stood out from the background. However, we cannot exclude that they did not *understand* the meaning associated to it. Independent of the cognitive reason, we state that the interface was in breach of the EU Digital Service Act and the Online Safety Bill, which explicitly will require easy access.

Foundational difficulties addressed the *clarity* of the report, the *efficiency* of the system and the *distress* induced by reporting such a hideous abuse. A summary of the results is provided in Table 4. It is evident that different interaction styles bring forward unique benefits and specific challenges. *Scriptum* allowed the user to create the clearest report, but was strongly disliked due to the embarrassment of having to describe the crime in writing. *Bonam* and *Malam* tended to be assessed in a very similar way, because the probability that the user identified the correct option was relatively low. Elaborating on these findings, the paper brings forward methodological guidance and a design contribution.

Table 4 Summary of usability results

Dimension	Variable	Ranking
Efficacy	Completion	Scriptum=Bonam=Malam
	Clarity	Scriptum>Bonam>Malam
Efficiency	Time	Scriptum<Bonam=Malam
	Click	Scriptum>Bonam>Malam
Satisfaction	Quantitative	Scriptum<Bonam=Malam
	Qualitative	Scriptum<Bonam>Malam

5.1 Methodological guidance

The recent legislative discussion about content moderation imposes a deep reflection on how to design reporting interfaces [20]. The paper enriches the generic and open-textured EU legal requirements of providing "easy to access and user-friendly mechanisms for notifying the presence of illegal content online" [26]. Grounding on the legislation surrounding secondary victimisation, we claim that usability studies should be enforced by law to demonstrate that the provider has put sufficient caution in designing effective and respectful interfaces. At this aim, the paper contributes with methodological guidance by providing an evaluation procedure (based on a medium-fidelity prototype) and a quality metric (based on a short questionnaire).

Researchers and practitioners could exploit the method to test existing and advanced interfaces according to the metric defined. The method proposes operational measures of usability in terms of efficacy (completion and clarity of the report), efficiency (clicks and time), and satisfaction where elements of distress and embarrassment are prevalent. In particular, to measure the psychological reaction to the act of reporting online gender based abuse, we propose a seven-item scale of user experience. The scale proved robust inter-item reliability and external validity when compared to the qualitative evaluation. This scale may be used in experimental studies or real life cases to assess the experience of the victim.

5.2 Design contribution

The paper represents an initial step in addressing legal complexity in technology design [34] with particular attention to improving the user experience in a context where only a nascent body of literature exists [42]. Reporting a crime online shall be precise and leave no room for misunderstanding. Our participants demonstrated to prefer a menu-based solution, but this requires the identification of clear items, which takes into account complex differences in platform jurisdiction. In this context, the vocabulary used in the menu-interface assumes foundational relevance. The option most often selected in *Malam* was *pornography and nudity*. However, pornography refers to certain adult content, legitimate, and for mass consumption [47] and nudity indicates something irrelevant from a criminal perspective. Non-consensual pornography was also often described as a kind of sexual harassment. Nevertheless, the concept of sexual harassment is more general and includes a variety of behaviours, not always relevant from a criminal perspective. Non-specific expressions will inevitably produce unsatisfactory, inexact, and incomplete reports.

Open text narratives appeared inadequate too, because the report was often incomplete or the conduct wrongly described, due to embarrassment and lack of legal knowledge. Only 23% of the participants using *Scriptum* included in their descriptions that the intimate images were shared without consent, a fact that hampers action from a legal viewpoint. In other words, our study highlighted how potential victims might face not only distress and embarrassment but also concrete difficulties in producing a narrative of the abuse with complete information. Platform designers should strive to find unambiguous triggers of immediate action, seeking for inspiration in the human-factors literature dealing with safety critical systems. Even at the cost of trading aesthetics for efficiency, we need standardised panic buttons.

The key point in improving moderation is the ability to distinguish between legitimate pornography and non-consensual pornography. The difference does not emerge directly from the images or videos themselves, nor it can be identifiable, because the same picture can be posted with or without consent. For this reason, the design of automatic reactive solutions should include the human-in-the loop, and we need respectful interfaces. We can imagine the application of an adaptive interface triggered by media recognition algorithms [53] that inspect reported material on the fly. If nudity is identified, a menu-based interface is displayed, with explicit reference to non-consensual pornography as a visible first level item. The user can then confirm and expand on it, but the embarrassment of the verbal description is offloaded on algorithms. The exact wording of the item shall follow the legal terminology applied by the jurisdiction of the platform, as inferred by the contact details reported in it. Further explanation and psychological support may also be introduced at this point.

5.3 Limitations and future research

With respect to the limitations and potential bias of this work, we are aware that the participants were not exposed to real crimes and therefore they were not victims. Post-trauma reaction should be further investigated involving law enforcement authorities, psychologists, and victims' associations. A future line of research in this direction could lead to case studies involving NGOs, policymakers, and governmental institutions. These studies could unveil a more realistic and vivid description of the victim's experience and how technology could improve it. Real case analysis is also fundamental from a legal perspective to evaluate that regulation, legislation, or terms of use introduced against non-consensual pornography are effective.

Another important limitation of the study is the small sample of young and educated adults. Despite taking a victim perspective, the studies did not consider the victim as a person with specific demographic characteristics, personality, and culture. Further research on gender and other intersectional aspects (e.g., race, ability) are paramount since women and girls are often the victims of non-consensual pornography [75]. As demonstrated by [2] women suffer higher victims blame and minority groups, especially sexual minorities, are most often the target of cyberbullying [46].

In this direction, our work aims at applying a victim-centred approach to the design of legal services and interfaces. There is an urgent need of new reporting mechanisms which can truly interface laypeople and the law. Victims need to be put in the situation to comprehend and improve their own legal situation and their role, as victims, in the justice system. This goal requires one to consider several factors including the design of effective and efficient reporting interfaces alongside the provision of emotional support to the victims [48], who are experiencing a violation of their human rights.

6 Conclusion

The paper has identified a dramatic situation as regards victims of non-consensual pornography who want to report the crime online. Not only is the abuse very difficult to be notified, but also the notification can expose the user to substantial distress. The analysis of real cases [41] suggests that the contested content can easily spread on several platforms. In this tremendous hypothesis, victims have to move through several interfaces, which are not only difficult to use (and often hidden) but likely to subject them to secondary victimisation, as evident in the *Scriptum* condition. Yet, these interfaces are currently implemented also on institutional websites of law enforcement agencies for official online reporting [18, 62, 63]. There are foundational challenges for legal and computer science research starting from the finding of this paper that people prefer menu-based selection to open-ended narratives because

> Its embarrassing to think that someone will read that. [P3, Scriptum].

Considering the prevalence of platforms which allow pornography in their terms of use and the complex legal framework applicable to digital platforms, the terminology and the structure of menu-based interfaces must be improved to support *clarity*, maintain *efficiency*, while minimising *distress*. This task pertains to HCI research. Instead, a pressing challenge for multimedia research is that of elaborating technical solutions to support self-reporting by offsetting the negative emotions associated to the description to a machine, as in the scenario previously described. We conclude by calling for increased liability on the providers to ensure that notification interfaces are not only available but can also be effectively used, in such a way that it feels safe and minimises distress. Legal research may contribute to this goal by connecting the EU Victims Rights Directive and the upcoming EU Digital service. HCI researchers shall develop user experience standards. Multimedia research may offer new sophisticated algorithms to help the user to describe offending images. In these challenging tasks, they can benefit from the metrics and the methodological guidance presented in this paper to avoid future victims will feel confused or unable to define what they have survived.

Appendix

Content Reporting Menus

Bonam

- Nudity
  - Adult Nudity
  - Sexually suggestive
  - Sexual Activity
  - Sexual exploration
  - Sexual services
  - Involves a child
  - Sharing private images
- Violence
  - Graphic violence
  - Death of severe injury
  - Violent threat
  - Animal Abuse
  - Something else
- Harassment
  - me
  - a friend
- Suicide or self-injury
- False information
  - Health
  - Politics
  - Social issue

Malam

Misleading information
Covid19
Other
Illegal activities and regulated goods
Promotion criminal activities
Sales or use of weapons
Drugs and controlled substances
Violent and graphic content
Suicide, self-harm, and dangerous acts
Suicide
Self-harm
Dangerous acts
Harassment or bullying
Me
Someone i know
Celebrity
Others
Minor safety
Underground delinquent behaviour
Child abuse
Inappropriate for minors
Intellectual propert infringement

    - Something else
- Spam
- Unauthorised sales
    - Drugs
    - Weapons
    - Endangered animals
    - Other animals
    - Something else
- Hate speech
    - Race or ethnicity
    - National origin
    - Religious affiliation
    - Social caste
    - Sexual orientation
    - Sex or gender identity
    - Disability or disease
    - Something else
- Terrorism
- Something else
    - Intellectual property
    - Fraud or scam
    - Mocking victims
    - Bullying
    - Child Abuse
    - Animal Abuse
    - Sexual activity
    - Suicide or self injury
    - Hate speech
    - Promoting drug use
    - Non-consensual intimite images
    - Sexual exploitation
    - Harassment
    - Unauthorised sales
    - Violence
    - Sharing private images

Dangerous organisation and individuals
Terrorism
Hate groups
Criminal groups
Frauds and scams
Animal cruelty
Hate speech
Pornography and nudity
Spam
Other

Funding and Competing Interests Open access funding provided by Libera Università di Bolzano within the CRUI-CARE Agreement. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no financial or proprietary interests in any material discussed in this article.

Data Availability The raw data that support the findings of this study are available in Zenodo with the identifier https://doi.org/10.5281/zenodo.6630392.

Declarations

Ethical Review This study has been carried out in the context of the *CREEP* project[3] studying the phenomenon of non-consensual distribution of intimate images of adults. Due to the sensitive topic, several

[3] https://creep.projects.unibz.it

activities related to the project required a through evaluation of ethical issues; therefore specific requests for approval were submitted to the Ethics Review Board of the Free University of Bozen-Bolzano.

Before conducting the study, a complete description of the procedure and evaluation methodology was submitted to the board. In particular, the documentation confirmed that:

- informed consent was obtained before the study, with both written and verbal explanation;
- the anonymity of participants was guaranteed and the data were managed according to the GDPR regulation;
- participants were not going to be exposed to explicit images;
- participants had the possibility to abandon the study at any stage and no data would be collected.

The Ethics Review Board approved the request in November 2021.

References

1. Aleksandra Kuczerawy (2019) From 'Notice and take down' to 'Notice and stay down': risks and safeguards for freedom of expression. In: Oxford handbook of online intermediary liability. Giancarlo Frosio
2. Attrill-Smith A, Wesson CJ, Chater ML, Weekes L (2021) Gender differences in videoed accounts of victim blaming for revenge porn for self-taken and stealth-taken sexually explicit images and videos. Cyberpsychology: J Psychosocial Res Cyberspace, vol 15(4). https://doi.org/10.5817/CP2021-4-3
3. Auxier B, Anderson M (2021) Social media use in 2021. Technical report, pew research center
4. (2022) Anonymous: code for the evaluation of reporting modalities zenodo. https://doi.org/10.5281/zenodo.5821243
5. Bar-Ziv S, Elkin-Koren N (2018) Behind the scenes of online copyright enforcement: empirical evidence on notice & takedown. Connecticut Law Rev 50(2):339–386. Online available at: https://opencommons.uconn.edu/law_review/402
6. Bartneck C, Brahnam S, Angeli AD, Pelachaud C (2008) Editorial: special section on misuse and abuse of interactive technologies. Interact Stud 9(3):397–401. https://doi.org/10.1075/is.9.3.01edi
7. Bellini R, Dell N, Whitty M, Bhattacharya D, Wall D, Briggs P (2020) Crime and/or punishment: joining the dots between crime, legality and HCI. In: Extended abstracts of the 2020 CHI conference on human factors in computing systems. ACM, pp 1–8. https://doi.org/10.1145/3334480.3375176
8. Bevan N, Carter J, Harker S (2015) ISO 9241-11 revised: what have we learnt about usability since 1998? In: Kurosu M (ed) Human-computer interaction: design and evaluation. Springer international publishing, vol 9169, pp 143–151. https://doi.org/10.1007/978-3-319-20901-2_13
9. Blackwell L, Dimond J, Schoenebeck S, Lampe C (2017) Classification and its consequences for online harassment: design insights from heartmob. Proc ACM Human-Comput Inter 1(CSCW):1–19. https://doi.org/10.1145/3134659
10. Bothamley S, Tully RJ (2017) Understanding revenge pornography: public perceptions of revenge pornography and victim blaming. J Aggression Conflict Peace Res 10(1):1–10. https://doi.org/10.1108/JACPR-09-2016-0253
11. Brahnam S, De Angeli A (2008) Special issue on the abuse and misuse of social agents. Interact Comput 20(3):287–291. https://doi.org/10.1016/j.intcom.2008.02.001
12. Brahnam S, De Angeli A (2012) Gender affordances of conversational agents. Interact Comput 24(3):139–153. https://doi.org/10.1016/j.intcom.2012.05.001
13. Bundeskriminalamt (2022) BKA - police crime statistics. Last Accessed on 09 Jan 2022
14. Caletti GM (2021) Can affirmative consent save Revenge Porn Laws? Lessons from the italian criminalization of non-consensual pornography. Virginia J Law Technol, vol 25(3). Online available at: https://www.vjolt.org/s/v25i3Caletti.pdf

15. Cooper K, Quayle E, Jonsson L, Svedin CG (2016) Adolescents and self-taken sexual images: a review of the literature. Comput Hum Behav 55:706–716. https://doi.org/10.1016/j.chb.2015.10.003
16. Cotten SR, Schuster AM, Seifert A (2021) Social media use and well-being among older adults. Current opinion in psychology. https://doi.org/10.1016/j.copsyc.2021.12.005
17. Crawford K, Gillespie T (2016) What is a flag for? Social media reporting tools and the vocabulary of complaint. New Media Society 18(3):410–428. https://doi.org/10.117/1461444814543163
18. (2022) Crimestoppersuk.org: crimestoppers. Speak up. Stay Safe. https://crimestoppers-uk.org/give-information/forms/generic-form?crimetypes=Other
19. De Angeli A, Brahnam S (2008) I hate you! disinhibition with virtual partners. Interact Comput 20(3):302–310. https://doi.org/10.1016/j.intcom.2008.02.004
20. De Angeli A, Falduti M, Menéndez-Blanco M, Tessaris S (2021) Reporting revenge porn: a preliminary expert analysis. In: Proceedings of the 14th biannual conference of the italian SIGCHI chapter, CHItaly '21, Bozen-Bolzano, Italy, and Online (Www), 11-13 July 2021, pp 31–35. https://doi.org/10.1145/3464385.3464739
21. De Angeli A, Hartmann J, Sutcliffe A (2009) The effect of brand on the evaluation of websites. In: Gross T, Gulliksen J, Kotzé P, Oestreicher L, Palanque P, Prates RO, Winckler M (eds) Human-computer interaction–INTERACT 2009. Springer, Berlin, Heidelberg, vol 5727, pp 638–651. https://doi.org/10.1007/978-3-642-03658-3_69
22. De Angeli A, Johnson GI, Coventry L (2001) The unfriendly user: second international conference on affective human factor design, Singapore, 27-29 June. In: Proceedings of the international conference on affective human factors design. Conference code: 2nd, pp 467–474
23. De Angeli A, Sutcliffe A, Hartmann J (2006) Interaction, usability and aesthetics: What influences users' preferences? In: Proceedings of the 6th conference on designing interactive systems. DIS '06. Association for computing machinery, pp 271–280. https://doi.org/10.1145/1142405.1142446
24. Drouin M, Vogel KN, Surbey A, Stills JR (2013) Let's talk about sexting, baby: computer-mediated sexual behaviors among young adults. Comput Hum Behav 29(5):25–30. https://doi.org/10.1016/j.chb.2012.12.030
25. Etikan İ, Musa SA, Alkassim R (2016) Comparison of convenience sampling and purposive sampling. Amer J Theo Appl Stat 5(1):1–4. https://doi.org/10.11648/J.AJTAS.20160501.11
26. (2000) European Commission: proposal for a regulation of the european parliament and of the council on a single market for digital services (digital services act) and amending directive 2000/31/EC
27. (2012) EU parliament and council: 2012/29/EU - the victims' rights directive, establishing minimum standards on the rights, support and protection of victims of crime, and replacing council framework decision 2001/220/JHA
28. (2017) European institute for gender equality (EIGE): cyber violence against women and girls. Last Accessed on 09 Jan 2022. https://doi.org/10.2839/876816
29. (2020) European parliament: online platforms' moderation of illegal content online. Law, practices and options for reform. Technical report EU
30. Fido D, Harper CA (2020) An introduction to image-based sexual abuse. In: Non-consensual image-based sexual offending: bridging legal and psychological perspectives. Springer international publishing, pp 1–26. 10.1007/978-3-030-59284-4_1
31. Freed D, Palmer J, Minchala DE, Levy K, Ristenpart T, Dell N (2017) Digital technologies and intimate partner violence: a qualitative analysis with multiple stakeholders. Proc of the ACM on Human-Comput Inter 1(CSCW):1–22. https://doi.org/10.1145/3134681
32. Freed D, Palmer J, Minchala D, Levy K, Ristenpart T, Dell N (2018) "A stalker's paradise": how intimate partner abusers exploit technology. In: Proc of the 2018 CHI conference on human factors in computing systems. Association for computing machinery, pp 1–13
33. Furlo N, Gleason J, Feun K, Zytko D (2021) Rethinking dating apps as sexual consent apps: a new use case for ai-mediated communication. In: Companion publication of the 2021 conference on computer supported cooperative work and social computing. CSCW '21. Association for computing machinery, pp 53–56. https://doi.org/10.1145/3462204.3481770
34. Goldstein D, Hill E, Lazar J, Siempelkamp A, Taylor A, Lepofsky D (2011) Increasing legal requirements for interface accessibility
35. (2022) Government bill: online safety bill - a bill to make provision for and in connection with the regulation by OFCOM of certain internet services; for and in connection with communications offences; and for connected purposes
36. Halder D, Jaishankar K (2011) Cyber gender harassment and secondary victimization: a comparative analysis of the united states, the UK, and India. Vict Offenders 6(4):386–398. https://doi.org/10.1080/15564886.2011.607402

37. Hartmann J, Sutcliffe A, Angeli AD (2008) Towards a theory of user judgment of aesthetics and user interface quality. ACM Trans Comput-Human Inter 15(4):15–11530. https://doi.org/10.1145/1460355.1460357
38. Hearn J, Hall M (2019) “This is my cheating ex”: gender and sexuality in revenge porn. Sexualities 22(5-6):860–882. Publisher: sage publications sage UK: London, England
39. Im J, Dimond J, Berton M, Lee U, Mustelier K, Ackerman MS, Gilbert E (2021) Yes: affirmative consent as a theoretical framework for understanding and imagining social platforms. In: CHI conference on human factors in computing systems. Association for computing machinery, pp 1–18
40. Katarya R, Lal A (2020) A study on combating emerging threat of deepfake weaponization. Proc 4th Int Conf IoT Social, Mobile, Anal Cloud (I-SMAC):485–490. https://doi.org/10.1109/I-SMAC49090.2020.9243588
41. Kleeman J (2018) The YouTube star who fought back against revenge porn – and won. The guardian
42. Kou Y, Gui X (2021) Flag and flaggability in automated moderation: the case of reporting toxic behavior in an online game community. In: Proceedings of the 2021 CHI conference on human factors in computing systems. ACM, pp 1–12. https://doi.org/10.1145/3411764.3445279
43. Kumar A, Bhavsar A, Verma R (2020) Detecting deepfakes with metric learning. In: 2020 8th International workshop on biometrics and forensics (IWBF), pp 1–6. https://doi.org/10.1109/IWBF49977.2020.9107962
44. Kumar A, Sachdeva N (2019) Cyberbullying detection on social multimedia using soft computing techniques: a meta-analysis. Multimed Tools Appl 78(17):23973–24010. https://doi.org/10.1007/s11042-019-7234-z
45. Liagre F, Verleysen C (2016) European crime prevention network (EUCPN) toolbox series. Preventing secondary victimization policies & practices. Technical report, prevention of and fight against crime programme of the european union european commission – directorate-general home affairs
46. Llorent VJ, Ortega-Ruiz R, Zych I (2016) Bullying and cyberbullying in minorities: are they more vulnerable than the majority group?. Frontiers Psychol, vol 7. https://doi.org/10.3389/fpsyg.2016.01507
47. Maddocks S (2019) Revenge porn: 5 important reasons why we should not call it by that name. GenderIt.org
48. Maeng W, Lee J (2022) Designing and evaluating a chatbot for survivors of image-based sexual abuse. In: CHI conference on human factors in computing systems. ACM, pp 1–21. 10.1145/3491102.3517629
49. Marganski A, Melander L (2018) Intimate partner violence victimization in the cyber and real world: examining the extent of cyber aggression experiences and its association with in-person dating violence. J Interpersonal Viol 33(7):1071–1095. https://doi.org/10.1177/0886260515614283. Publisher: SAGE publications inc
50. Marwick AE (2017) Scandal or sex crime? Gendered privacy and the celebrity nude photo leaks. Ethics Inf Technol 19(3):177–191. https://doi.org/10.1007/s10676-017-9431-7
51. McCann W, Pedneault A, Stohr MK, Hemmens C (2018) Upskirting: a statutory analysis of legislative responses to video voyeurism 10 years down the road. Criminal Justice Rev 43(4):399–418. https://doi.org/10.1177/0734016817741342
52. McGlynn C, Rackley E (2017) Image-based sexual abuse. Oxford J Legal Studies 37(3):534–561. https://doi.org/10.1093/ojls/gqw033. _eprint: https://academic.oup.com/ojls/article-pdf/37/3/534/32374499/gqw033.pdf
53. Mohanty M, Zhang M, Russello G (2019) A photo forensics-based prototype to combat revenge porn. In: Proc of the 4th IEEE conf on multimedia information processing and retrieval (MIPR), pp 5–8. https://doi.org/10.1109/MIPR.2019.00009
54. Montada L (1994) Injustice in harm and loss. Soc Justice Res 7(1):5–28. https://doi.org/10.1007/BF02333820
55. North Yorkshire Police (2018) Fire and crime commissioner: suffering in silence: why revenge porn victims are afraid and unwilling to come forward because of a fear they’ll be named and shamed – and why that needs to change. Last Accessed on 09 Jan 2022
56. Nurse JRC (2018) Cybercrime and you: how criminals attack and the human factors that they seek to exploit. The Oxford Handbook of Cyberpsychology, pp 662–690. https://doi.org/10.1093/oxfordhb/9780198812746.013.35
57. O’Connell A (2020) Image rights and image wrongs: image-based sexual abuse and online takedown. J Intell Property Law Pract 15(1):55–66. https://doi.org/10.1093/jiplp/jpz150
58. O’Malley RL, Holt KM (2020) Cyber sextortion: an exploratory analysis of different perpetrators engaging in a similar crime. J Interpersonal Violence. https://doi.org/10.1177/0886260520909186
59. Obar JA, Oeldorf-Hirsch A (2020) The biggest lie on the internet: ignoring the privacy policies and terms of service policies of social networking services. Inf Commun Society 23(1):128–147. https://doi.org/10.1080/1369118X.2018.1486870

60. Orth U (2002) Secondary victimization of crime victims by criminal proceedings. Soc Justice Res 15(4):313–325. https://doi.org/10.1023/A:1021210323461
61. Pandya A, Lodha P (2021) Social connectedness, excessive screen time during COVID-19 and mental health: a review of current evidence. Frontiers Human Dynamics 3:45. https://doi.org/10.3389/fhumd.2021.684137
62. Polizei Berlin (2022) Internetwache Polizei Berlin. https://www.internetwache-polizei-berlin.de/index_start.html
63. Polizia di Stato (2022) Segnala Online. https://www.commissariatodips.it/segnalazioni/segnala-online/index.html
64. Polizia di Stato (2022) I dati 2021 della Polizia postale. Last Accessed 9 Jan 2022
65. Powell A, Henry N, Flynn A, Scott AJ (2019) Image-based sexual abuse: the extent, nature, and predictors of perpetration in a community sample of australian residents. Comput Hum Behav 92:393–402
66. Price M, Verhulst S (2000) The concept of self-regulation and the internet. Protecting our children on the internet: towards a new culture of responsibility 1(1):133–198
67. Salehzadeh Niksirat K, Anthoine-Milhomme E, Randin S, Huguenin K, Cherubini M (2021) I thought you were okay : participatory design with young adults to fight multiparty privacy conflicts in online social networks Designing interactive systems conference 2021. ACM, pp 104–124. https://doi.org/10.1145/3461778.3462040
68. Sambasivan N, Batool A, Ahmed N, Matthews T, Thomas K, Gaytán-Lugo LS, Nemer D, Bursztein E, Churchill E, Consolvo S (2019) “They don’t leave us alone anywhere we go”: gender and digital abuse in south asia. In: CHI conference on human factors in computing systems. Association for computing machinery, pp 1–14
69. Smith JA (ed) (2015) Qualitative Psychology: A Practical Guide to Research Methods, 3rd edn. SAGE, London
70. Strohmayer A, Clamen J, Laing M (2019) Technologies for social justice: lessons from sex workers on the front lines. In: CHI conference on human factors in computing systems. CHI ’19. Association for computing machinery, pp 1–14. https://doi.org/10.1145/3290605.3300882
71. Takhar P (2018) A proposal for a notice-and-takedown process for revenge porn. Harvard Journal Of Law & Technology Digest. Online available at: https://jolt.law.harvard.edu/digest/a-proposal-for-a-notice-and-takedown-process-for-revenge-porn
72. Terry K (2021) Pandemic fuels 329% rise in revenge porn offences in london: met police records 1,185 cases of private snaps and videos illegally shared online in last year - with victims as young as TEN. Last Accessed 9 Jan 2022
73. Vallina P, Feal Á, Gamba J, Vallina-Rodriguez N, Anta AF (2019) Tales from the porn: a comprehensive privacy analysis of the web porn ecosystem. In: Proc of the internet measurement conference. ACM, pp 245–258. https://doi.org/10.1145/3355369.3355583
74. Van de Weijer SGA, Leukfeldt R, Bernasco W (2019) Determinants of reporting cybercrime: a comparison between identity theft, consumer fraud, and hacking. Eur J Criminol 16(4):486–508. https://doi.org/10.1177/1477370818773610
75. Web Foundation (2021) How online gender-based violence affects the safety of young women and girls. https://webfoundation.org/2021/03/how-online-gender-based-violence-affects-the-safety-of-young-women-and-girls/
76. Yan J (2017) Novel security and privacy perspectives of camera fingerprints (transcript of discussion). In: Anderson J, Matyáš V, Christianson B, Stajano F (eds) Security protocols, XXIV. Springer international publishing, pp 96–102
77. Yeh C-Y, Chen H-W, Tsai S-L, Wang S-D (2020) Disrupting image-translation-based deepfake algorithms with adversarial attacks. In: Proc of the IEEE/CVF winter conf on applications of computer vision workshops, pp 53–62

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.