U.PORTO

2º CICLO

MESTRADO EM MEDICIAL LEGAL

Online Grooming and Online Safety: How Do Web Platforms Detect, Prevent and Moderate

Joana Resende

M

2022

U.PORTO

Joana Resende. Online Grooming and Online Safety: How Do Web Platforms Detect, Prevent and Moderate

M.ICBAS **2022**

Online Grooming and Online Safety: How Do Web Platforms Detect, Prevent and Moderate

Joana Resende

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

Joana Sofia Melo Resende

Online Grooming and Online Safety: How Do Web Platforms Detect, Prevent and Moderate

Dissertação de Candidatura ao grau de Mestre em Medicina Legal submetida ao Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto.

Orientador – Doutor Carlos Peixoto

Categoria – Professor assistente convidado

Afiliação – Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto. (originalmente)

Co-Orientador – Doutora Maria José Carneiro de Sousa Pinto da Costa

Categoria – Professor Associado

Afiliação – Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto. (originalmente)

Abstract

Life is increasingly lived in less physical environments. Now, more than ever, the status quo is populated by information of all kinds, shapes and sizes.

There's barely an *online* anymore. In a sense it is all just beginning to mesh together into a new and very differentiated mode of existence. We don't simply go online anymore, we *live* online, we *are* online. And when online, we live everywhere and anywhere (provided your internet connection is stable enough).

The internet is the entrance to an ultimate plane of existence, feeding (off) the broken vessel that is the mind and permeating our lives, coalescing into a nearly unrecognisable, smooth blend of sometimes multiple *I*'s. It's easier to find people to identify with inside a screen. For better or worse, under the cloaked idea of near anonymity, people find their peers and assemble in undeniable validation.

While there's an increasing interest in this theme and the scientific research focusing solely on online sexual grooming is growing, there is a little number of the studies that provide a direct insight on how internet platforms, particularly those whose main goal is to connect individuals, tackle this problem.

The aim of this dissertation is to study some of the most egregious behaviours online and how moderation teams tackle these conducts and groups on online platforms. The goal of this thesis is to provide some insights into the ways with which these teams identify, deal with and prevent this type of online behaviour and material.

Keywords: moderation; social media; cybercrime; pedophilic behaviours; grooming; online safety

Index

List of Abbreviations

AI — Artificial Intelligence

CSAM — Child Sexual Abuse Material

E2EE — End-to-end Encryption

GDPR — General Data Protection Regulation

IWF — Internet Watch Foundation

ML — Machine Learning

NCMEC — National Center for Missing and Exploited Children

OG — Online Grooming

OSTIA — Online Safety Tech Industry Association

SbD — Safety by Design

SG-CSAM — Self-generated Child Sexual Abuse Material

U.S.C. — United States Code

U.S.C. — United States Code

VPN — Virtual Private Network

CP — Child Pornography

MAA — Minor Attracted Adult

I

Internet-Mediated Existence

The main goal of this short thesis is to explore the practice of online grooming and CSAM repeated offenders and provide some insight into how virtual platforms tackle this increasingly pressing issue, including a cursory review into its legal status. In the first portion, I'll cover definitions and data related to the internet and mainly social media platforms. Next, I'll broach some of the most proiminent theories and data concerning internet-facilitated sexual offences. Lastly, in the third and final portion of this analysis, I'll focus on providing insight regarding the way in which online platforms detect the presence of minor abuse related content, mitigate its risks and attempt to prevent the behaviour from taking place within their "bounds". Briefly, this thesis is a meta-analysis of the problem from a content moderator's perspective.

Ever since the internet made its public debut in the early to mid 90s, the way in which civilization connects to knowledge, each other and opportunities faced a change in paradigm. As the years went by and more and more resources and content made its way to the ever expanding ocean of the virtual, it became increasingly easier to find what one is looking for, and sometimes not looking for. Its presence is so marked that some authors believe the equal access to the digital world (along with the right to disengage from technologically mediated opportunities) must be contemplated in the UN's Universal Declaration of Human Rights (Risse, 2021; Woodroofe, 2020). However, as a human made and human-occupied tool, the internet is plagued with human made and humanlike challenges. While the so-called Web 1.0 was much like a catalogue of information ready for consumption, its second iteration, Web 2.0, made us all not only consumers but the product. The gathering of data we willingly shared with internet companies is exploited for commercial gain when sold to advertisers. Now, the surfacing of Web 3.0 may come as a direct reaction to manipulation attempts as it proposes the creation of a decentralised internet where, among other ideas, gathering of data by big tech is hindered and cryptocurrencies and similar exchange mediums roam freely. If detecting and tackling deviant, abusive behaviours online is quite challenging in a Web 2.0 world, Web 3.0 will certainly complicate it further.

A Virtual Playground

Just what exactly is a social network? What climate provides fertile ground for its blossoming and popularisation? Obar & Wildman (2015), proposed two fundamental conditions in the

emergence of these new ways of socialisation that we call social networks: the development of Web 2.0 as the natively interaction features it brought with itself and the decrease of digital content storing prices. Together, these aspects allowed platforms to offer spaces entirely dedicated to interaction to its users. However, reducing social networks into a mere way of facilitating communication, similar to phones or letters, is disingenuous. Even less so, to pose that a social network is only the result of technological developments associated with the internet is reductive and rash. Hence, I have reflected on the nature of a social network. Indubitably, a series of aspects emerge as fundamental for us to consider a platform as a social network: (1) the existence of a personal, almost non-transferable profile that serves a digital function equivalent to state identity documents, a *meta-personna*, (2) the possibility of connecting users and (3) sharing information and content either publicly (e.g. the Wall on Facebook) or privately (via direct messages). The latter is perhaps the primordial and defining characteristic of a social network. User-generated content is the lifeblood and the profile is the backbone of the social network (Obar & Wildman, 2015; Boyd & Ellison, 2007). Sharing and visibility of a specific user's connections is a crucial component of the social network (Boyd & Ellison, 2007), allowing users of a given network to find profiles of others of potential interest to them. Let us now consider the following: if social networks allow individuals to establish relationships with their acquaintances and peers, it is therefore not surprising that people with similar interests cross paths in the virtual world. Like minded individuals construct their own niches, finding camaraderie among themselves. This is a two edged sword: one may find acceptance in a community, helping each other cope with the life's inherent pains and troubles but said acceptance may also fuel less than socially acceptable, if not illegal, behaviour. And while perhaps, this consideration may appear bleak, it does not make it less true. Given the scope of what a social network is, several issues related to a variety of topics arise: protection of privacy, freedom of expression, intellectual property and, most pertinently for this analysis, the protection of minors (Obar & Wildman, 2015). From this less than ideal standpoint, the internet can be seen as an assembly of deviant behaviours. In 1998, Cooper coined the term *Triple A Engine* — access, affordability and anonymity — to underline the three main characteristics that not only make the internet so appealing but which also surfaces the dangers associated with it**.** The internet non-infrequently presents itself as an unsupervised (parent or authority-free), free entry and opportunity-ridden playground, that requires little to no information for anyone that wishes to visit. Despite the era of "digital natives" to whom there's barely little to no distinction between online and offline living, there appears to be a propensity for startling attitudes or behaviours which may present as "outside of the realms of our normal characteristics" (Whittle et al, 2013).

Digital Natives

In 2001, Prensky popularised the term *Digital Natives* to refer to generations of people to whom the use of the internet and consequent fluency is second nature. As per the latest estimates, more than half of the world population (61.8%) uses the internet (Kemp, 2022). 67.1% of the population has at least one mobile communication device (5.29 billion individuals) and the number of internet users increased by 4.8% between October 2020 and October 2021. The estimate points to 4.88 billion people with internet access, of which more than 93% have an account on at least one social network. Among the 20 most visited sites we find Google, Youtube, Facebook (1st, 2nd and 3rd respectively) but also Pornhub, XVideos, XNXX, the last three dedicated to sharing pornographic material. Most access time takes place on mobile devices (with an average of more than 3 hours of use), corresponding to 90.9% of total accesses. Not surprisingly, social networks emerge as the most visited type of website (with about 2h30min of daily access) led by the Facebook platform, followed by the use of chat platforms. Although most users have an account for each of the different platforms, as a general rule, the different applications tend to attract different age groups. For example, among young people aged between 16 and 24, Instagram emerges as the preferred platform. The Snapchat application, on the other hand, is particularly popular among the youngest (13-17 years), showing a decreasing trend the older the user. Furthermore, some of these applications are particularly appealing to younger people precisely because they are less popular among older age groups, reducing the likelihood that children will cross paths with their parents (Obar & Wildman, 2015).

In Portugal, data[1] suggest that 80% of the population makes daily use of the internet (Eurostat), values that rise to 87% if expanded to the entire European community. EU data show that 1 in 3 internet users is a minor (European Commission, 2019). 2020 data for the UK estimates that almost all children aged between 5 and 15 have been online at some point, which is not entirely unexpected considering that many of these children attended online classes. via tools such as Zoom, resulting from the measures implemented in response to the Covid-19 pandemic (Ofcom, 2021). Perhaps more surprising is the fact that from the age of 5 onwards, the number of children who have a device with internet access rises to more than 50%, with a considerable percentage of children and young people using social networks (55%). Perhaps the most worrying consideration is that 40% of children aged between 8 and 10 use social media, a percentage that grows considerably with increasing age with pre-teens and teenagers — 87% between 12 and 15 year olds (Ofcom, 2021). These show a remarkable while not unexpected increase when comparing with data from 2010, which showed 38% and 77% usage of social networks by younger and older

[1] Estimated data for the year 2021

teenagers respectively (Whittle et al, 2013). A particular concern arises here, as children frequently expose themselves online, which makes it easier for potential predators to identify available victims. More often than not, children below the age of 13 do not disclose their actual age on the account, otherwise they would not be able to launch it[2]. However, this does not mean that they refrain from posting their information (photos, activities, and so on) online. Children often create pages to follow others of their interest (friends, family members, actors, influencers, games and game communities, etc). Knowing that particular individuals, communities or apps are more likely to attract young people, will make it easier for online predators to find potential victims. This is something that permeates all platforms. So how do platforms fight this type of behaviour? How are these prevented, detected and dealt with? We will explore this in section III of this thesis.

II

An Assembly Of Deviant Behaviour

Making use of the internet to share and engage in predatory pedophilic behaviours is nothing new, nor will this thesis claim such. However, we can not simply discard the fact that the increased and spread usage of the tool and the proliferation of social media platforms played an important role in the evolution and propagation of this behaviour but so did awareness on the matter. In this section, we will cover some of the most common, often egregious, pedophilic behaviours taking place in an internet-mediated environment. We'll also consider the theories proposed to understand why individuals seek sexual interactions with minors.

Theories On Sexual Offences

In 1984, Finkelhor described the first multifactorial theory (Ward et al, 2006) related to child sexual abuse, coined the *Precondition Model*. The model accounts for "*the psychological needs and motives of the offender, situational and contextual variables, parenting practices and social attitudes towards children and sex*" (Ward et al, 2006) and essentially suggests the presence of four preconditions to sexual offending (Craven et al, 2006). The first one is the very motivation to sexually abuse a child which is associated with three different motives: emotional congruence (the individual's needs and the child who can

[2] Legal consequence established by the US' COPPA federal law of 1998. On the other hand, and while largest and most tech companies are North American and based in the USA, the European Commission appears to be piggybacking on this law as it does not establish a general minimum age at which an individual can open an account, at the discretion of each country. Portugal has no recommendation in this regard, as can be seen with mere access to Google's Help Center.

meet those), sexual arousal (deviant sexual arousal to children which is theorized by the author may have been caused by a string of "*maladaptive early learning experiences*" (Ward et al, 2006) and blockage (where the "*sexual needs of the offender may not be met by appropriate adults*" (Craven et al, 2006) (Ward et al, 2006; Craven et al, 2006). The second and third conditions focus on the individual's internal and external inhibitors. As for internal inhibitors, fleeting states of disinhibition such as inebriation or drug consumption, or less momentary ones (psychosis, senility, stress, etc) may present fertile ground for an individual's lowering inhibitions (Ward et al, 2006). Ward et al further discuss how extrinsic factors such as *hostile masculinity* (Seto, 2017), social tolerance towards peadophilic tendencies or the availability of CSAM on the internet may which may "*function as cognitive distortions and cause men to interpret potential sexual situations with children in self-serving ways*" (Ward et al, 2006). The third precondition refers to overcoming external constraints that prevent the abuse, such as parents or caretakers. Lastly, with the first three conditions met, the individual's sexually motives to abuse must find ways to overcome the child's resistance. There are several strategies employed to reach this objective such as manipulation tactics, rapport building and desensitization to sexual content (Ward et al, 2006, Whittle et al, 2013; Kleijn & Bogaerts, 2020), which will we'll ellaborate further on later in this paper. These last two conditions present what can and has been described as grooming (Craven et al, 2006).

While Finkelhor's Precondition Model was most likely the first to propose a theory as to why (and how) adults engage in sexual victimizatrion of young individuals (Ward et al, 2006), it certainly wasn't the last. In 1990, Marshall and Barbaree proposed an *integrated theory* which suggests how "*the presence of vulnerabilities which develop as a result of adverse early developmental experiences, leave offenders unprepared to deal with the surge of hormones at puberty, and unable to understand the emotional world*" (Craven et al, 2006). In short, the theory proposes that hormone influx and the presence of traits such as "*low self-esteem, a poor coping style and inadequate interpersonal skills*" (Ward et al, 2006) increase the probability of young males acting in sexual deviant ways. Craven et al point out that the theory suggests the deviant act occurs in the aftermath of the fusion between sex interest and aggression drive (Craven et al, 2006). This entails the presence of aggressive abuse (in which the individual is not capable of discerning between sexual arousal and aggressive states such as anger) which doesn't account for abuse processes such as grooming, which often rely on less imediate goal-achiving tactics (Craven et al, 2006; Ward et al, 2006).

In 1992, Hall and Hirschman publish their *Quadripartite Model*, which was initially developed as a theory of rape but later applied to child sexual abuse (Craven et al, 2006). The authors propose that the sexual abuse occurs due to the presence of four vulnerabity

factors and an opportunity (Craven et al, 2006). The authors claim that the sexual (physical) arousal to children, cognitive distortions, affective dyscontrol and personality issues in conjuction with an oportunity could result in child abuse, however, how all the factors come together is unclear and the theory relies on the idea that for the abuse to manisfest one or a combination of the vulnerabilities must exceed a threshold (Ward et al, 2006; Craven et al, 2006). Furthermore, Craven et al argue that because "*sexual grooming is not an impulsive act*" the circumstances that lead to overpassing the limit, would have to remain the same for an esxtended period of time (Craven et al, 2006).

Building and in an attempt to "knit" together (Ward et al, 2006) Finkelhor's, Marshall and Barbaree's, as well as Hall and Hirschman's models, Ward and Siegert proposed a *Pathway Model* in 2002. The model is based on the presence of dysfunctions in physiological mechanisms which can be summarizes into emotional dysregulation (for example, inability to recognize one's emotional state), interpersonal competence (low social skills or difficulty establishing intimacy), cognition distortions (maladaptive beliefs) and sexual preferences (Ward et al, 2006; Craven et al, 2006). The theory attempts to account for the act while considering the heterogeneity among offenders, as the dysfunctions will differ in levels of severity between individuals and will result in an offence when conjugated with a sexual need (Ward et al, 2006; Craven et al, 2006). However, the model fails to reflect on the actual process of offending which Craven et al, (2006) argue is fundamental in the development of a behaviour model.

In 2012, Lanning distinguishes and typified two main groups of (cyber) sex offenders: *situational offenders* and *preferential offenders.* The first is classified as less intelligent, mostly from low-income demographics, who act impulsively and opportunistically. They may molest children, not due to a preference but because these are easier, more available victims. These offenders are also associated with higher rates of criminal offences. Within this type of offender, Lanning defined three subgroups of cyber offenders: (a) "*normal*" *teenager/adult* (curious or prurient users), (b) *morally indiscriminate offenders* (presenting a variety of violent criminal sex offences) and (c) *profiteer* (financially motivated offenders). Conversely, *preferential offenders* do display specific tendencies, are more intelligent and from higher social-economic groups. Their behaviour is, according to Lanning, "*primarily fantasy-driven*", compulsive, and more discriminate in their patterns and techniques. Among this group, Lanning named three subtypes: (a) *paedophiles*, (b) *diverse offenders* (nearly indiscriminate offenders, displaying interests in several paraphilic content), (c) *latent offenders* (previously offence-free individuals but who already held potentially illegal preferences) and provided identifiable characteristics of *preferential offenders* such as the collection of themed sexual content, rationalization of sex interests or display long term patterns of behaviour.

Ward & Keenan (1999, mentioned in Ward et al, 2006) identified five cognitive schemas (Elliott & Beech, 2009), cognitive distortions or implicit theories in child sexual abusers: (1) *children as sexual beings* — children share the same needs and wants with adults; (2) *nature of harm* — children are being harmed in the process; (3) *entitlement* — the offender is superior to the child; (4) *dangerous world* — people are not to be trusted*;* and (5) *uncountability* — the world, rather than the individual, is uncontrollable and their actions are not their fault. Taylor & Quayle (2003, mentioned in Elliott & Beech, 2009) categorised four cognitive distortions found in internet offenders: *Category 1*: in which offenders justify the use of CP since those are only pictures; *Category 2*: those that normalise the behaviour as so many others do it; *Category 3*: objectification of the images through collecting; and *Category 4*: interaction with images and even real children in a way to cooperate with the community.

While understanding the motivation to sexually offend is tantalizingly important, the actual process of how the offence takes place may provide authorities and other regulatory agents (online platforms) with insights to tackle online crimes. If a clearer understanding of the overall process is reached, positioning roadblocks and checkpoints along the chain of events may help reduce the number and success of the offences. Ward, Louden, Hudson & Marshall 1995's *Descriptive Model of Offence Chain* identifies nine stages in the offending process. Most of the stages feature subcategories which will affect either positively or negatively the upcoming stage. The first stage is preoccupied with background factors or how the individual perceives their own lifestyle and circumstances, extending the notions of the self (Ward et al, 2013). Stage two represents distal planning, which may feature explicit or implicit planning or the presence of an opportunity (Ward et al, 2013; Craven et al, 2006). The contact with the victim comes in the third stage. The fourth stage consists of cognitive restructuring where the presence of sexual arousal and/or cognitive distortions may have a positive or negative effect, followed by proximal planning (fifth stage). The sexual offence (stage six) will reflect the focus chosen by the offender (self-focused, victim needs' focused or mutual focus). The next stage (seven) will reflect a second cognitive restructuring whose negative or positive effect will determine if the perpetrator will re-offend and establish future resolutions (stage eight). The ninth and final stage (background factors) will reflect these decisions of whether the individual will persist in the behaviour or choose to avoid it (Ward et al, 2013; Craven et al, 2006).

More recently, Seto (2017) proposed a *Motivation-Facilitation Model* which identifies traits and states that coupled with opportunity, may lead to sexual offences. Seto establishes motivation as the first needed step and identifies three sexual motivations: paraphilias (described by the American Psychiatric Association (American Psychiatric Association [APA], 2013, p. 685) as "*intense, persistent sexual interest in other than sexual interest in genital*

stimulation or preparatory fondling with phenotypically normal, physically mature, consenting human partners"), high sex drive and intense mating effort (defined as the "*energy devoted toward pursuing and obtaining new partners*" (Seto et al, 2019). The presence of one of the above mentioned motivations or a combination thereof does not equal offence, as facilitating circumstances must take place in conjugation. Seto describes these facilitators as "*factors that overcome any trait or state inhibitions against acting upon motivations*" which can be exemplified by impulsivity, recklessness (self-regulation problems), misogynistic and hostile behaviour towards women and children, disinhibitors (alcohol or illegal substances) and anger, stress or depressive states (negative affect). However, Seto argues an opportunity such as the presence of a vulnerable victim in the absence of guardians must take place for the offence to come to pass. The internet allows for the individuals to seek themselves said opportunity (McGrath & Casey, 2002).

Seto (2012), argues that *pedophilia can be defined as a sexual attraction to prepubescent children, as indicated by persistent and recurrent sexual thoughts, fantasies, urges, arousal, or behavior*". Pedophilia is described similarly by the American Psychiatric Association (American Psychiatric Association [APA], 2013, p. 697) as a "*recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving sexual activity with a prepubescent child or children (generally age 13 years or younger)*". The second criteria refers to the existance of an action "*on these sexual urges, or the sexual urges or fantasies cause marked distress or interpersonal difficulty*". The APA also establishes the minimum age of 16 and the presence of at least a 5 year difference between the individual and the child of interest. Some authors distinguish between pedophilia and hebephilia in which the latter specifically refers to individuals presenting sexual interest in pubescent children (Bailey, Hsu & Bernhard, 2016; Bailey, Bernhard & Hsu, 2016; Jones et al, 2020) and the term pedohebephilia can be observed in the literature to refer sexual interest in children and young adolescents.

While establishing an accurate percentage on prevalence is of particular difficulty due to the negatively charged social consequences (Seto, 2012), research estimates that between 3% and 9% of men (far less common in women) experience sexual attraction to children (American Psychiatric Association, 2013; Seto, 2012; Mitchell, Bravo & Galupo, 2017). It has been argued that paedophilia can be seen as a sexual orientation, displaying comparable characteristics as any other orientation such as (a) age of onset (in pubescent years), (b) sexual and romantic behaviour (emotional congruence with children), and (c) stability over time (interest remains throughout the individual's lifespan) (Seto, 2011; Tozdan & Briken, 2015; Bailey, Hsu & Bernhard, 2016). While an individual may exhibit pedophilic tendencies, they may still retain non-deviant interests and studies suggest child abuse in which the victim is a child is more frequently correlated with more normative experiences

and a lower recidivism rate (Mitchell, Bravo & Galupo, 2017). It is important to underline that while an adult engaging in sexual activity with a physically mature minor (such as a 15-year-old) is a crime, it does not qualify as pedohebephilia and paedophiles may exhibit attraction towards consenting adults (Seto, 2012). It is believed that paedophilia closely interacts with antisociality, a history of abuse and neurodevelopmental perturbations (American Psychiatric Association, 2013; Seto, 2012).

I believe we must always be careful not to equate pedophilic tendencies to pedophilic actions. Many of those diagnosed and/or displaying pedophilic tendencies never do engage in illegal acts, their behaviours bordering but never actually crossing the line into legally punishable behaviour (Seto, 2012; Bailey, Bernhard & Hsu, 2016). These people may be referred to as a minor attracted persons or virtuous paedophiles that make use of a myriad of coping mechanisms to stay offence-free (Jones et al, 2020). As expected, the internet is also populated by websites, such as forums and private groups, that welcome individuals displaying these tendencies, providing validation and camaraderie that may or may not prevent future illegal acts. There is such a group of people that define themselves as *minor-attracted adult* (MAA)*, child lovers, boy lovers, and girl lovers*, among other terms, that attempt to mitigate the stigma associated with this paraphilia. Holt, Blevins and Burkert (2010), point out that the social sphere of those displaying pedophilic tendencies "*is shaped by four interrelated normative orders including marginalisation, sexuality, law and security*". In these forums, one can and is often expected to share their preferences and coping mechanisms (be it to refrain from engaging in *real-life activity* or to justify their actions), as for marginalised communities such as these, the sharing of some degree of information is crucial to guarantee trust and foment interaction. Here individuals can also find and offer information pertaining to the applicable laws and how to keep themselves safe from detection. This *virtual self-disclosure* allows them to enter a community of like-minded people that both strengthens and supports involvement in deviant behaviour (Holt, Blevins and Burkert, 2010).

The identification of pedophilic individuals typically occurs after the perpetrator has been caught attempting or after incurring in child abuse or child abuse-related activities such as the possessing, production and/or distribution of child pornographic material. While the details in its definition may differ, child pornography can be described as any sexually explicit material featuring a person under the age of 18.

On offence supporting beliefs

Several authors have identified the consumption of child pornography as a offence supporting belief (or cognitive distortions), fueling the offender's fantasies and "constructing"

(Rimer, 2019) an idea of a child as both a sexual being and a sexual object (Bartels & Merdian, 2015; Paquette & Cortoni, 2020). It can be postulated that child pornography's legitimizing nature is two-fold: on one hand children in these materials are often smiling and seemingly content which fuels the belief that no harm comes from the behaviour and it becomes easier for the offenders to disassociate with the child who becomes “not-real” (Rimer, 2019), on the other hand, this content was produced by someone which legitimizes the fantasies as others have done it as well (Prat et al, 2020). Among Paquette and Cortini’s (2020) sample of CSAM offenders, 73.33% viewed children as sexual beings, capable of consent. This value decreased to 66% when it came to online groomers. The authors found that, in general, for these online offenders, acts of rape or murder are significantly more severe than child pornography. For them, the virtual is not real and their actions and urges are something outside of their control. In fact, this perception of uncontrollability has been observed and proposed previously, specifically by Ward (2000) who identified it as one of the 5 main cognitive distortions related to child abuse. Alongside the belief in the inability to control one’s own actions, the nature of the harm, and the idea of children as sexual objects, Ward also refers to entitlement, as the offenders are often observed to believe they can do as they please, which Paquette and Cortini verified in their sample; and the dangerous world implicit theory. For offenders, relationships with children pose less danger than adults (less abusive and less rejecting) and because the world, in general, is threatening, these individuals feel the need to assert their own dominance (Ward 2000). Paquette and Cortini also identified the belief of children as potential sexual partners due to emotional congruence and the perception of an egalitarian relationship. This was also observed by Setisteban et al’s (2018) study in which the offenders not only blamed the victim, claiming they have been deceived about their age and intentions but asserted the minor played an active role in the interaction as a *provocateur*.

While cognitive distortions help justify and minimise the offender’s actions, many offenders find coping strategies to deal with their desires and urges. Jones et al’s (2020) study of a forum geared towards pedophilic individuals identified a set of seven strategies to refrain from engaging in legally punished behaviors: (1) having contact rules such as only engaging with children when the contact is initiated by the minor; (2) mentally preparing for potentially risky situations by imagining scenarios and organising their lives in an effort to reduce the likelihood of encountering children; (3) finding hobbies to shift the attention away from their sexual preocupations; (4) avoiding children-geared areas altogether as well as dishinibitory substances such as alchool looking to minimize opportunity; (5) removing the temptation by having internet access rules; (6) considering the consequences of the actions, be it from the victim’s perspective or the legal consequences. Similarly, and in support of this strategy, Geradt et al (2018) have found that frequent contact with children relates to fewer

offences supporting beliefs. Finally, (7) seeking legal outlets such as animated material of child characters or pornographic content featuring young looking actors or masturbating to deal with sexual arousal. The authors point out that the usage of legal pornographic material featuring age ambiguous or animated characters may be a potential maladaptive strategy.

On CSAM

In the late 1990s, in an attempt to forensically qualify the degree of severity of the material the COPINE project (*COmbating Paedophile Information Networks in Europe*) was launched and introduced the COPINE scale (Quayle, 2008). The scale sets 10 levels of severity, ranging from non-explicit to sadistic imagery, as outlined below:

Level 1: Indicative	Non-erotic and non-sexualized children in underwear or other types of minimal clothing.
Level 2: Nudist	Naked or semi-naked imagery of children in nudist, non-sexualized environments.
Level 3: Erotica	Surreptitious photos eliciting a sexualized gaze of naked, semi-naked children.
Level 4: Posing	Deliberate posing of children (naked or otherwise).
Level 5: Erotic Posing	Children (naked or otherwise) in provocative or suggestive poses
Level 6: Explicit Erotic Posing	Photos emphasising children's genitals (the child may not naked or not)
Level 7: Explicit Sexual Activity	Explicit sexual activity involving children.
Level 8: Assault	Explicit sexual activity involving children and an adult.
Level 9: Gross Assault	Explicit and obscene sexual activity involving children and adults.
Level 10: Sadistic/Bestiality	Presence of violent interactions and intercourse/sexual activity involving a child and an animal.

Building upon this, Krone (2004) developed a *typology of online child pornography offending* which qualified child pornography offenders into 9 different categories considering their level of involvelment with the material: *Browser* (accidental user that decides to keep the offending

content), *Private Fantasy* (someone who fantasizes having intercourse with a chikd but has not committed an offfence, writes or creates digital images to satisfy the fantasy), *Trawler* (someone who actively seeks child pornographic material, maybe an omivorous or curious user), *Non-secure collector* (purchases, downloads or exchanges CP from non-secure sources), *Secure collector* (a more careful collector, emplyoing tacts to avoid detection), *Groomer* (cultivates a relationship with child online which may or may not evolve into physical contact, may use CP to desensatize the victim and normalize the behaviour), *Physical abuser* (offline abuse of a child, relationship may have been established through online grooming), *Producer* (offender who records their own abuse or those of others) and *Distributor* (this individual may or may not have interest in children but distributes the material).

Usage of child pornography (CP), child sexual abuse material (CSAM) or child exploitation material (CEM) is seen as a fairly straightforward indicator of sexual interest in children (Seto, 2012; Seto et al, 2010; Babchishin et al, 2014; McCarthy, 2010; Seto & Eke, 2017), although these may not have translated into a contact offence. It's also of paramount importance to consider that child sexual abuse is not exclusive to sexualized imagery of children or contact offences but also encapsulates any activity featuring a sexually-charged power imbalance between an adult and a child such as voyeurism, exhibitionism and the like (Wild et al, 2019).

Prat & Jonas (2012) argue that the majority of online sexual abusers are CSAM consumers and describe two major types of such individuals: (a) those who access the files in absence of an opportunity to offend and (b) those who use the online world to express their socially unaccepted fantasies. Many authors (Burgess and Hartman, 2005[3]; Elliot and Beech, 2009[4]; Krone, 2004[5]) have classified these individuals into different types of CSAM consumers, among them Beech, Elliott, Birgden & Findlater (2008) who identified four major categories: (1) "*curious and impulsive users*", (2) those "*accessing and trading images to fuel their sexual interests*", (3) "*contact offenders who also use child pornography*", and (4) "*those who disseminate images for non-sexual reasons*". Merdian et al (2013) describe 5 reasons for the usage of CSAM: (1) identify and establish contact with like minded individuals, (2) to engage in sexually charged interactions with underage persons, (3) to harass, bully, or threaten said victims with, (4) to locate vulnerable victims and (5) to promote the content and child trafficking, fueled by four distinct motivations: (1) "*paedophilic motivation*", (2) "*general deviant sexual interest*", (3) "*financial motivation*" and (4) "*other*" such as curious users. While individuals can display a combination of some or all of the above identified reasons,

[3] *travellers* and *traffickers*
[4] *direct victimizers* and *commercial exploiters*
[5] *physical abusers*, *producers* and *distributors*

many offenders are consumer-only offenders. Quayle & Taylor (2002) conducted a study which focused on the reasons offered by offenders themselves as to why they engage in this type of deviant behaviour. The authors reported that the most relayed reason was to achieve sexual arousal, followed by the urge to collect and categorizing the pictures (one of the offenders normalized the activity by comparing it to the collection of baseball cards while other desensatized it by equating it to trophies). Quayle & Taylor also found that the content would be used as a way to facilitate social interactions with individuals with similar interests which "*enabled social cohesion*" and can be argued further fuels cognitive distortions particularly those concerning feelings of alienation. Furthermore, it will mostly likely entail the sharing of knowledge on how to remain undetected. The material was also described as a way to avoid real life, finding more fulfilling relationships within the members of the community. Lastly, offenders reported that the usage of CSAM was a therapy, a way to refrain from engaging in "hands on" offences. Similarly, Beech et al (2008), in a comprehensive criminological review, outlined four broad typologies of internet CSAM consumers: (1) sporadic, curious or impulsive users, (2) those who access and/or trade the material to cater to their paedophilic tendencies, (3) individuals who use the internet to expand a pattern of offline offences and seek potential victims or share content they have produced and (4) others to whom use doesn't appear to provide any sexual gratification. Merdian et al (2013) proposed a classification of child pornography offenders into subgroups: *fantasy-driven* and *contact-driven offenders*, in which the former while engaging in trading of pornography material and virtual content with children do not seek real world content which is the reverse in the latter.

On the offenders

In the literature, three types of offenders have been described: (1) contact (offline) offenders, (2) internet (online) offenders and (3) mixed offenders (Elliott et al, 2013; Seto & Hanson, 2011). Authors such as Briggs et al (2010) and Babchishin et al (2010) have considered how the internet provided fertile ground for the emergence of a new type of offender: the online offender. These can be summarily described as sex offenders who use the internet in their crimes (usage of CP, solicitation, indecent exposure, etc) and who share both similarities and differences with contact offenders (Babchishin et al, 2010). Babchishin et al (2010) postulated that the internet, in its vast array of potential targets and largely unregulated spaces, could lead otherwise offence-free individuals to commit sexual crimes. In their meta-analysis of several other studies, the authors found that most of the individuals convicted of online offences were caucasian and younger than offline offenders (38 years vs. 46 years of age) which may be explained by an unequal internet availability distribution. This

has been supported by Winters et al (2017)'s findings. Furthermore, the analysis denoted differences in victim empathy (less found in offline offenders), sexual deviancy (higher rates in online offenders), in the presence of cognitive distortions and emotional congruence (in which contact offenders rated higher). This is also in line with Elliott, Beech & Mandeville-Norden's 2012's study findings, where the authors noted that online offenders were less likely to hold beliefs that characterise the interaction with the child as harmful. Online offenders also appear to be less antisocial than contact offenders but estimates suggest one in eight child pornography offenders has had offline sexual contact with a child (Babchishin, Hanson & VanZuylen, 2013). The authors also drew comparisons between online sex offenders and the general population and discovered these individuals presented a significantly higher history of physical and sexual abuse victimization. In addition, these offenders were more likely to be unemployed (despite the fact that the authors found no statistical difference in levels of education), never married or unmarried at the time of the offence and presenting a history of substance abuse. Babchishin, Hanson & VanZuylen performed a similar study in 2013, expanding the comparison to mixed offenders, i.e. offenders with both online and offline offences, and found further distinctions between child pornography offenders and mixed offenders. Mixed offenders showed greater pedohebephiliac tendencies than online-only and offline-only offenders and are also more likely to have access to children than online-only offenders. This has also been corroborated in McManus et al. study from 2015. Additionally, mixed offenders show similar antisociality markers when compared with offline offenders but present higher empathy deficits. Findings from other research suggest mixed offenders are more likely to cross over and/or re-offend than online-only offenders (Babchishin, Hanson & VanZuylen, 2014; Hirschtritt et al, 2019; Soldino, 2019). Alexy et al (2005) on an analysis of media reports regarding sexual offences against children, distinguished between *traders* (those who collect and trade/traffic CSAM), *travellers* (who contact and groom children online with the goal of committing an offline offence) and *combination traders-travellers* (individuals who both trade/collect CSAM and travel to engage in sexual intercourse with a child).

On Online Grooming: *Intimacy at arm's length*[6]

As previously referred, CP is frequently used by offenders (mainly *contact-driven* mixed offenders) in the grooming process as a way to desensitise the victims and normalise the behaviour proposed by the perpetrator (Merdian et al, 2013; Beech, Elliott, Birgden & Findlater, 2008). Building upon Merdian et al (2013)'s *fantasy-driven* and *contact-driven* offenders, Briggs et al, (2011) analysed a sample of 51 individuals convicted of

[6] *In* Whittle et al, 2013.

internet-initiated sexual offences targetting children following a police sting operation and found sufficient evidence to distinguish between two subtypes: *fantasy-driven* and *contact-driven* chat room offenders. While the former appears to find gratification in cyber-contacts with the victims which include suggestive language, masturbation and/or cybersex; the latter aims to meet and develop (employing grooming tactics) relationships with children which will later evolve into in-real-life contact. Here, the internet becomes a medium that expands the individual's opportunities to offend and the authors identified a specific offender type — the chatroom offender. DeHart et al (2016) further differentiated these *internet-based* offenders into *cybersex-only offenders*, *schedulers*, *cybersex/schedulers* and *buyers*. The research's results note that *cybersex-only offenders* are mainly white men who tend to expose themselves to and request explicit material from the victim. Similarly, *cybersex/schedulers,* also white in general, seeked explicit web-mediated sexual interactions with the victims (self-exposure and masturbation) but, out of the four categories, were more likely to present pedohebephilic and incest interest. This subgroup would discuss the scheduling of an event, specifying time and date but were also the most likely to cancel it. *Schedulers*, while primarily white did figure non-white individuals. In contrast to the former, these offenders were less likely to expose themselves and their focus was on quickly scheduling an in-person contact. Lastly, *buyers*, like *schedulers* were more diverse, but scheduling appeared to be the main focus of the digital interaction while inquiring about the prices associated with offline engagements. Furthermore, the study's results also alerted to the fact that the initially seemingly naive interactions may result in sexualized interchanges quickly (sometimes in less than 10 minutes) which has been observed in other studies as well (such as Kleijn & Bogaerts 2020's research into patterns of engagement in webcam child sex tourism offenders). Kloess et al (2015) found several strategies by which web-based offenders engage and achieve compliance with victims and identified four strategies: (a) *directness in initiating online sexual activity* (explicit language, sending to or requesting explicit pictures from the victims), (b) *pursuing sexual information* (offenders inquire victims about their preferences, experiences and pratices), (c) *the next step* or an attempt to lure the victim into a physical meeting and (d) *fantasy rehearsal* (where the perpetrator shares their fantasy with the victim). Additionally, the study also discovered manipulative strategies such as the usage of pornographic material (self generated or otherwise) to supplement sexual stimulation and security seeking snippets such as asking the victim if they were keeping their relationship secret.

But what exactly is grooming? Craven et al (2007), describe online grooming as "*a process by which a person prepares a child, significant adults and the environment for the abuse of this child. Specific goals include gaining access to the child, gaining the child's compliance and maintaining the child's secrecy to avoid disclosure. This process serves to*

strengthen the offender's abusive pattern, as it may be used as a means of justifying or denying their actions." Williams, Elliot and Beech (2013) expanded this definition to "*a process by which an individual prepares the child and their environment for abuse to take place, including gaining access to the child, creating compliance and trust, and ensuring secrecy to avoid disclosure*". Webster et al (2012) posited the process of grooming is the result of an empirical process, in which the offender employs tactics that have been successful for them in the past.

Kloess et al (2014), while describing the process of grooming, identified three types of sexual grooming: (a) *self-grooming*, (b) *grooming of the environment and significant others* and (c) *grooming of the child*. *Self-grooming* refers to the mechanisms the perpetrating individual employs to justify their actions and wants, such as cognitive distortions. *Grooming of the environment and significant others* refers to methods used by the offender to gain the trust of the child's guardians (parents, teachers, etc) to avoid detection and further the (continuous) access to the child. As an example, the authors of this study and others suggest groomers may choose single-parent families or absent-parent families as primary targets, as these would be, arguably, more susceptible to exploitation. Lastly, *grooming of the child* relates to all the tactics exerted to influence the child into an abusive sexual relationship. These have been summarised by O'Connell (2003) as follows:

(a) *Friendship forming stage* — the time period in which the perpetrator attempts to know the child, they may ask about their interests, hobbies and the like. While there is an empirical expectation for the offender to pose as a non-adult individual, in fact, research has found that that may not always be the case (Briggs et al, (2011); O'Connell, 2003; Marcum, 2007) nor does it substantially alter the chances of offline encounters (Bergen et al, 2014). Additionally, Mitchell et al 2005's research showed that internet-mediated grooming is also used by family members or acquaintances of the victims, which represented 39% of the sample.

(b) *Relationship forming stage* — the offender tries to deepen the interaction, showing concern and interest about the child's wellbeing, creating an "*illusion of being the child's best friend*" (O'Connell, 2003). The offender may also use flattery to gain the child's trust (Black et al, 2015; Kloess et al, 2015).

(c) *Risk assessment stage* — marked by attempts by the perpetrating individual to understand their chances of being caught, of the interaction being disclosed. P.J. Black et al's study (2015) into the language employed when grooming, revealed that this stage may come earlier than postulated by O'Connell as the words "*mother, father, worry, nervous and home*" were used more frequently in the first segments of the analysed transcripts and at the beginning of each interaction.

(d) *Exclusivity stage* — the idea of trust usually is introduced at this stage, with the offenders showing themselves as worthy of the child's trust;

(e) *Sexual stage* — following the intimacy established by the previous stage, offenders may start to engage the child in more personal, sexually charged themes. O'Connell (2003) points out that this is the stage where the most differences between tactics can be observed. The theme may be introduced in a gentle manner or more aggressively. This is also the stage at which pornography is introduced as a way to normalise the interaction or the offender may choose to enact their fantasies with the victim. Alternatively, the individual may coerce, threaten and manipulate the child to engage in the activity through strategies such as bribery, deception or a false sense of involvement as found by de Santisteban et al (2018) and Seymour-Smith & Kloess' (2021) studies.

(f) *Concluding stage* — here, the offender may employ what the author coins "*damage limitation*" tactics such as praising or encouraging the child. Alternatively, the individual may simply choose to "*hit and run*" and cease contact with the victim abruptly.

Similarly, Williams, Elliot and Beech (2013) identified three themes in the process: (a) *rapport building* (O'Connell's *friendship* and *relationship forming* and *exclusivity stages*), (b) *sexual content* (in which the perpetrator introduces and later escalates the use of sexually charge themes and terms) and (c) *assessment* (where the groomer attempts to gauge the level of trust that be deposit in the child and their chances of sucess and detection). In another research, N. Lorenzo-Dus et al (2016) identified three phases of online grooming: (a) *access* (initial contact), (b) *approach* and (c) *entrapment* (the ultimate aim being to meet the child offline). The later authors also determined that online groomers display idiosyncratic features which can be summarised as a (1) extended use of explicit sexual solicitation, which it's postulated to be a desensitisation tactic, (2) rapport and trust building strategies and (3) marked preocupation in establishing victim's compliance. Prior to engaging in any sort of contact with the victim, however, offenders set out to scan for possible victims. In 2007, Malesky's research into online predatory behaviour found that a significant portion of the sample (80%) scoured youth geared chat rooms to meet minors and nearly half of the individuals reviewed the minor's profile to screen potential victims. The presence of any mention of sexual activity or openess to discuss the theme, a submissive or "needy" looking minor and "*young sounding*" name appear to motivate contact by the offenders.

Specifically analysing online groomers, Webster et al (European Online Grooming Project, 2012) identified three major types of offenders, taking into consideration several

dimensions such as records of sexual offences, changes to their identity when establishing contact with children online, type of cognitive distortions emplyoed, etc:

(a) *intimacy-seeking*: individuals with no previous convictions for sexual offences, that seek a *consenting relationship* with a child and are less likely to change they identity online. These individuals do not possess indecent images of children or contact like-minded offenders online. Additionally, Webster et al found that these individuals present more emotional congruence with children, maintain long-term and frequent contact with the victim and is less aggressive in their introductions of sexual themes and seeks offline contact to solidify the relationship.

(b) *adaptable style*: this group tends to have had convictions for sexual offences against children and show offence-enabling beliefs (seeing the victim as a mature and capable person). Some may collect CSAM but do not appear to have contact with other similar individuals online in general. Offline contact is not established as a way to further a relationship as opposed to the previously mentioned group. Whether these individuals alter their identity online, will depend on the victim's reaction and may have several identities simultaneously. The same applies to the cadency and explicitness of contact but these are generally short and straightforward which may be a result of the scanning process as these individuals seek young people with sexual screen names and/or photos. Webster et al found the individuals in this group may use blackmailing to further/maintain the relationship with the youth. The outcome of the interaction may remain online (cyber-sex) or evolve to offline contact.

(c) *hyper-sexualized*: hyper-sexualized individuals are characterised by marked sexual interest in children and possess vast CSAM and extreme adult pornography collections. Webster et al pose the pronounced interest in children and CP may be a result of saturation from adult pornographic content. These offenders adopt integrally different identities and are the fastest in escalating the interaction to a sexual dimension. Offensive supportive beliefs tend to dehumanise the victim and do not seek to develop any sort of relationship.

Additionally, the authors detected the presence of six defining features of online grooming which can be summarised as follows: *vulnerability* on the offender's end (characterised by situational and interpersonal relationship factors that trigger the individuals into online offending, such as losing their job or a partner)*, scanning* of the victims (looking for potential victims in online spaces frequently geared towards or used by young people then scanning their profiles for children that match their needs such as sexualized screen names or fitting the offender's appraisal type)*, identity* chosen to be used in the process (some offenders alter their identity online, while others do not; the persona used in these interactions is, as per the authors, a signal for the type of offender)*, contact* style (mode,

cadency, style and timing differ significantly, some offenders may seek a long-term interaction while others are quite explicit and aggressive from the start making use of either chat, webcam or a combination of both)*, intensity* of the images, language and incentives shared with the victim (aiming to desensitised the child to sexual activity or normalizing the behaviour) and *outcome* of the interaction (which may result in offline contact).

Winters et al (2017)'s examination of 100 chatroom transcriptions found that, on average, online solicitation offenders/groomers introduced the theme of an offline meeting three days into the interaction. However, many of the perpetrators chose a direct approach and initially a sexualized interaction within the first 30 minutes.

On the victims

Several studies have also focused on the potential victim characteristics. While, as observed by Calvete et al (2020), most adolescents are at a low risk of victimisation and resilient to advances, authors have identified personal, interpersonal and eco-social idiosyncrasies frequently observed in victims of online exploitation. Webster et al (2012) differentiated at-risk adolescents into two groups: vulnerable victims and risk-taking victims. Among the former, gender and age account for individual vulnerabilities, where girls between the ages of 13 and 17 are more likely to be victimized than cisgender, heterosexual boys (Whittle et al, 2013; Munro, 2011; Winters et al, 2017), while gay, queer or questioning young men are more likely to be victimized (Whittle et al, 2013). Munro (2011) also observed that African-american girls are at a higher risk than caucasian young females. Adolescents with self-esteem and confidence issues, displaying loneliness and attention and affection needs (Webster et al, 2012) and/or presenting a disability are more likely to be victimised (Whittle et al, 2013; Munro, 2011; Winters et al, 2017). Difficult home lives (single-parent families, alcoholism, violent and dysfunctional parent-offspring relationships) also play a role in the presence of these susceptibilities, as teens seek comfort and belong in the online environment (Whittle et al, 2013; Webster et al, 2012). Whittle et al, 2013 postulated cultural vulnerabilities such as abuse supportive beliefs also present an added risk. On the other hand, risk-taking teenagers are more disinhibited, outgoing and confident and may or may not have faced offline abuse (Webster et al, 2012). Suler (2004)'s online disinhibition effect also shines some light into the beliefs and further adds to this behaviour, such as the concept of dissociative anonymity or invisibility. Risk-taking victims are more likely to engage in sexting with people they have met online (Munro, 2011; Calvete et al, 2020; Machimbarrena et al, 2018; Schoeps et al, 2020). This behaviour may lead to a quicker furthering of the sexualized interaction as offenders may use the sexual media as a way to entrap victims into compliance with threats of exposure (*sextortion*) as observed by

Seymour-Smith & Kloess' 2021 study. These characteristics offer insights into the offenders as well and, as suggested by Webster et al (2012), intimacy seeking offenders prefer vulnerable victims, while the hyper-sexualized offender will more likely seek risk-taking victims.

CSAM and OG in numbers

In this section, I'll briefly cover the statistics regarding CSAM and OG for the year 2021. As mentioned in the introductory portion of this thesis, access to the internet increased from 52% in the year 2000 to 88% in 2018, and it is safe to assume it has since leaped in percentage in the last four years. It's not erroneous to assume that most people in the world have access to the internet, particularly since all that is needed is a small portable electronic device and a bit of pocket change invested in access. Hence, the risk for increase in abusive and illegal practices has also increased. In fact, the European Commission Migration and Home Affairs has divulge 85 million images and videos featuring CSAM have been reported by internet companies worldwide in 2021 (European Commission, 2022)[7].

Greene-Colozzi el al (2020)'s study into perceptions of grooming by young people found that 25% of the participants engaged in online conversations with adults as a minor and 47% characterised the interaction as flirtatious and, in some cases, lead to offline sexual contact. Furthermore, 17% of the participants reported being a victim of online solicitation when underage and the overall results indicate that 23% of the sample was a victim of online sexual grooming.

Thorn's 2018 analysis points to an upward trend towards more egregious content (in the COPINE scale) being shared, where the most severe material (younger children, sadistic/violent abuse) usually involved a familial offender, whereas CSAM featuring pubescent children is much less likely to feature a family member. The report also found that content depicting an infant or toddler was the most likely to be actively traded. A 2020 study, also performed by Thorn, and specifically focusing on self-generated CSAM (SG-CSAM) found that 17% of children shared their own material (such as *nudes*) including 15% of children between 9 to 10 years of age, with 21% agreeing that it is normal to do so. LGBTQ+ children appear once again as more vulnerable, as 32% of the responders have shared SG-CSAM. 50% of the children who have shared CSAM admitted they have shared a nude image or video of someone they have not met in real life and 41% shared the content with someone they believed to be over the age of 18.

An Interpol report from 2020 denoted a significant increase in the sharing of CSAM in P2P networks and it also indicates a stark rise of self-generated CSAM (SG-CSAM) and the

[7] Note that these numbers do not reflect unique images or videos but the actual number of reports.

viralization of said content. This trend has of course been observed in the year 2021, as per Internet Watch Foundation's report, which marked an increase from 44% of the overall CSAM reports received in 2021 being SGCSAM to 72% in 2022. The Internet Watch Foundation's report features an analysis of the total number of reports reached from a variety of sources (hotlines, proactive search, authorities, etc). In 2021, the organisation reviewed over 300,000 reports, 69.8% of which featured actual child exploitation material or advertisement thereof (a 64% increase from 2020). The data suggests girls continued to be the most targeted gender (96.5%), however, more concerning is the increasingly younger age of the victims, as the majority were estimated to be between 11 to 13 years of age (68%) followed by 7 to 10 (23%), 3 to 6 (6%) and 14-15 (2%). The report also presents an in depth analysis of the webpages where the content was found, with image hosting websites and file hosting websites as the most common; 96% of the content was found on *free-to-use* services, 72% hosted in Europe (mainly in the Netherlands (42%)) and 17% in North America. Curiously, the report also shares a brief analysis of abuse perpetrated by female offenders, in which cases 49% of the content featured a male victim most commonly between the ages of 7 to 10. The number of recognizably female offenders in the material remains, however, strikingly low.

Additionally, a 2021 survey of 1200 minors conducted by Thorn and with a particular focus on Online Grooming, found that it is increasingly more common for minors to develop meaningful relationships with individuals they have met online, in fact, 1 in 3 minors considers someone that they have met online as a close friend, connections which are frequently established and maintained through shared interests (gaming, *fandom* subcultures[8], etc). However, the data suggested that, in general terms, minors tend to be wearier and less inclined to connect to and develop relationships with unfamiliar adults and lean more towards individuals they perceive to be of similar age. That said, 82% have reported interacting with adults they did not know in real life, indeed, 32% of teens (13 to 17 years old) connected with someone they believed to be between the ages of 20 and 29. More concerningly, 63% of the 9 to 12 year old minors who answered the survey reported interactions with adults online. It becomes even more alarming when considering that, in the sample, the data showed that 9 to 12-year-olds were far more likely, when compared with teenagers, to consider adults (who they perceived to be) over 30 as "close friends" rather than mere strangers or acquaintances. When considering the LGBTQ+ community, in particular, Thorn found that minors within this group were generally more comfortable interacting with online-only adult contacts than non-LGBTQ+ minors. The report also analyzes the perceived risks of online interactions, observing that 54% of minors consider

[8]As per Cambridge's online dictionary a fandom is *a group of fans of someone or something, especially very enthusiastic ones*

online grooming as a common experience with 2 in 5 minors reporting having been approached by an adult who they believe was trying to "*befriend and manipulate them*". Despite understanding the presence of a risk, 16% of 9 to 12 years old have had romantic or sexual conversations with an online-only contact and 13% have shared nude images of themselves with someone they knew exclusively online. The report also identified that 40% of minors had received a "cold solicitation" for explicit material online, with 1 in 3 9 to 12 year old boys reporting this type of interaction. 52% of the surveyed individuals have reported being asked to move from the public chat to a private conversation on a different platform and the same number of minors reported having experienced an uncomfortable conversation with an adult they perceived to be 30 or older. However, minors were far more likely to simply *ghost*[9] or block the user than they were to report the behaviour, which in 39% of the cases was due to a shift in the conversation (with someone perceived to be 18 or older) in which it became sexually charged. Interestingly, when asked where they had received information about the online risks and perils particularly concerning online grooming, the top two sources were parents/guardians and an online community they were part of which sheds some light on the role in-app moderation can play in guaranteeing a safe online place to children.

A 2021 Europol trend report also denotes a marked increase in online grooming related activity in social media and online gaming platforms and has identified SG-CSAM too as a key threat.

This data proves what many in the industry already knew: internet mediated crimes are becoming increasingly more commonplace. Furthermore, not only is the number of people being victimised rising exponentially but the age at the time of abuse appears to be decreasing. We can postulate that the ever-increasing access to the internet and acquisition of smartphones and/or tablets, exposes younger and younger children to the dangers of predators. However, access to the internet and the opportunities it presents should not be kept from children because back actors lurk in the shadows. Mainstream platforms are progressively more dedicated to the protection of children in the online environment but many, if not most, of the efforts to fight online child exploitation, remain unknown. In the next section, I'll attempt to shine some light on the tools, processes and people involved in the detection, mitigation, resolution and prevention of online child sexual abuse. It is not surprising that the majority of the material is found on websites with little to no moderation present. Electronic service providers play an important role in detecting, reporting and removing CSAM online. Some of these companies have been supporting the fight against online child sexual exploitation by applying specific technologies to proactively detect CSAM

[9] Cambridge's online dictionary describes *ghosting* as *a way of ending a relationship with someone suddenly by stopping all communication with them*

in their services. Detected material is then removed and referred to child protection NGOs and law enforcement agencies for analysis and investigation.

Legal considerations

The International Centre for Missing and Exploited Children (ICMEC), an international non-governmental organisation funded in 1999 focusing exclusively on the protection of children, performs a recurrent analysis of the legislation landscape concerning CSAM across the world. In the latest (9th) edition, the report indicates only 11 countries meet the 5 criteria the organisation established as necessary to fight child abuse: these being (1) the existence of a specific law penalizing CSAM, (2) providing a definition of child sexual abuse material, (3) featuring criminalization of technology-facilitated CSAM related offences, (4) and criminalizing the mere possession of said material as well as (5) requiring Internet Service Providers (ISPs) report the suspected to CSAM to law enforcement agencies or similar. While the US, Canada, France and India (only to name a few) do have legislation encompassing the 5 criteria, the vast majority of the countries do not. 71 countries meet the first four criteria and 79 (one of each Portugal) do not. 35 countries do not have any laws specifically criminalizing pornography. Per the report, while Portuguese Law does have specific legislation in place to fight child pornography, no concrete definition of the material exists nor is there a requirement for ISPs to report potentially illegal material. The same deficiency is found in Slovenia, Poland, Luxembourg and Indonesia legislation, just to name a few.

The 2007's Council of Europe Convention on Protection of Children against Sexual Exploitation and Sexual Abuse (also known as the Lanzarote Convention, effective 2010) promotes the criminalization of child pornography-related activities from producing to procuring child abuse material[10]. It also urges the presence of a definition of what constitutes child pornography (Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse, 2007, Article 20 (2)). In 2011, the Directive 2011/92/EU[11], adopted by the European Parliament and the Council of the European Union, expanded the definition of child pornography to "*(i) any material that visually depicts a child engaged in real*

[10] Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse (CETS 201), Article 20 (1) *Each Party shall take the necessary legislative or other measures to ensure that the following intentional conduct, when committed without right, is criminalised: a. producing child pornography; b. offering or making available child pornography; c. distributing or transmitting child pornography; d. procuring child pornography for oneself or for another person; e. possessing child pornography; f. knowingly obtaining access, through information and communication technologies, to child pornography.* (last visited Oct 9th, 2022)

[11] Online version; https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32011L0093 Last visited Oct 9th, 2022

or simulated sexually explicit conduct; (ii) any depiction of the sexual organs of a child for primarily sexual purposes; (iii) any material that visually depicts any person appearing to be a child engaged in real or simulated sexually explicit conduct or any depiction of the sexual organs of any person appearing to be a child, for primarily sexual purposes; or (iv) realistic images of a child engaged in sexually explicit conduct or realistic images of the sexual organs of a child, for primarily sexual purposes". In Portuguese law, child pornography is specifically criminalized under the article 176º of the Portuguese Penal Code, but with a vague definition of the abuse ("[…] *all the material that, for sexual purposes, represents minors involved in sexually explicit behaviours, real or simulated, or features any representation of their sexual organs or any other part of their body*"[12]). The production, distribution (or intention to), possession, use (or incitement to use) are prohibited under the same article. Grooming is also criminalised in the Portuguese Penal Code under the articles 171.º (c)[13] and 176.º-A[14] and is defined as the enticement by someone over the age of 18 that via information/communication technologies, grooms a minor aiming at the practice of any of the sexual acts.

In the US, a country where a vast number of internet companies are headquartered and whose Federal Laws these are first and foremost liable to, child pornographic material is criminalized under Title 18 of the United States Code (18 U.S.C. § 2251- Sexual Exploitation of Children; 18 U.S.C. § 2252- Certain activities relating to material involving the sexual exploitation of minors (Possession, distribution and receipt of child pornography); 18 U.S.C. § 2252A- certain activities relating to material constituting or containing child pornography). 18 U.S.C. § 2256 defines sexually explicit conduct as any real or simulated (i) sexual intercourse, including genital-genital, oral-genital, anal-genital, or oral-anal, whether between persons of the same or opposite sex; (ii)bestiality; (iii)masturbation; (iv)sadistic or masochistic abuse; or (v)lascivious exhibition of the anus, genitals, or pubic area of any person[15]. 18 U.S. Code § 2422(b) criminalises grooming[16].

---

[12] DECRETO-LEI N.º 48/95, DE 15 DE MARÇO; Artigo 176.º Pornografia de menores: 8 - Para efeitos do presente artigo, considera-se pornográfico todo o material que, com fins sexuais, represente menores envolvidos em comportamentos sexualmente explícitos, reais ou simulados, ou contenha qualquer representação dos seus órgãos sexuais ou de outra parte do seu corpo.

[13] *Id.* at Artigo 171.º Abuso sexual de crianças: c) Aliciar menor de 14 anos a assistir a abusos sexuais ou a atividades sexuais;

[14] *Id.* at Artigo 176.º Aliciamento de menores para fins sexuais: 1 - Quem, sendo maior, por meio de tecnologias de informação e de comunicação, aliciar menor, para encontro visando a prática de quaisquer dos atos compreendidos nos n.os 1 e 2 do artigo 171.º e nas alíneas a), b) e c) do n.º 1 do artigo anterior, é punido com pena de prisão até 1 ano.

[15] 18 U.S.C. § 2256 Online Version, https://www.govinfo.gov/content/pkg/USCODE-2015-title18/html/USCODE-2015-title18-partI-chap110.htm (Last visited Oct 9th, 2022)

[16] *Id* at Whoever, using the mail or any facility or means of interstate or foreign commerce, or within the special maritime and territorial jurisdiction of the United States knowingly persuades, induces,

III

Fighting the Good Fight

Platforms whose main goal is the hosting of user-generated content are posed with innumerable risks, and while its presence is far from new, Trust & Safety as a defined organisation within a company is still barely out of infancy. But what exactly is a Trust and Safety organisation? Mark Little, Kinzen[17]'s CEO and founder defines Trust and Safety as "*all those people in those big tech companies that are there to protect our conversations and our communities from [the kind of] toxic content and abuse*". Succinctly, we may define Trust and Safety as a policy enforcement group and, in fact, these teams are frequently compared to law enforcement agencies. Tech companies are not ethereal organisations detached and freed from the constraints of the analogue world. While the nature of its products may seem intangible at times (particularly if we think about companies such as Meta, Twitter, and TikTok), a virtual space in a corner of the internet, technology companies are, just like any other, subject to local, federal and national laws. A company's policies must abide by the laws of the countries where they are based in and, in some cases, operate. The Trust and Safety team's role is, as the name suggests, two-fold. It aims to ensure those law-based policies are followed, guaranteeing the safety of the company (by reducing the exposure to risk) and of its users (from potential bad actors and/or abuse within the platform) while building rapport and trust to improve engagement and retention. A Trust and Safety organisation is usually composed of two distinct regulatory bodies: a fraud team and a content moderation team. While the former focuses on (ideally) preventing fraud, the latter prioritises the evaluation of the content shared by the users on the platform. The bigger the platform, the larger the risk and, therefore, the greater the need for protective resources. Tech giants such as Youtube, Meta, or TikTok have significantly large moderation teams, distributed across several time zones in an effort to ensure extreme content is hosted on the platform for the shortest amount of time possible. Hence, it is expected that these first responders are frequently faced with exceptionally graphic and often egregious material.

Fraud and moderation teams are, in essence, virtual policy enforcement bodies, as a decentralised authority, particularly because these teams are generally specific to each

---

entices, or coerces any individual who has not attained the age of 18 years, to engage in prostitution or any sexual activity for which any person can be charged with a criminal offense, or attempts to do so, shall be fined under this title and imprisoned not less than 10 years or for life.

[17]Ireland-based content moderation tech company

company and its values[18]. They occupy a fringe place where they are simultaneously authorities and not, more than community moderators but less than recognized officials. Law enforced by non-law enforcers. Moderation teams, in particular, perform a unique role. The members of this group are tasked with analysing a piece of content and, often very quickly[19], expected to make a judgement call, based on the company's policy, as to whether or not the post, tweet, or video, is in violation of said policy. If so the material is taken down[20], if not it remains *live*. It shouldn't come as a surprise that moderation work is mostly reactive and that, despite the policies, the warnings, the in-app education, bad actors, edgy teens or simply distracted users find ways to engage in abusive conduct. Moderation protects all users from encountering sensitive, shocking, illegal content such as is the case for CSAM and OG.

The needle in the haystack

As mentioned above, moderation work is more often than not, reactive, similarly to what happens in the *non-virtual* space. How does potentially violative material reach a company's moderation team? In two ways: user reports and computer-mediated processes (information retrieval and artificial intelligence (AI)). The former is fairly straightforward: users submit a report for the content they are offended by and/or believe to be in violation of the policies. Often, users are also asked to choose the appropriate reasons so as to facilitate the identification of the issue and routing of the report[21]. Computer-mediated detection, however, is the largest violation-feeding funnel. For instance, a simple information retrieval tool known as Bag of Words (BoW for short) can be used as the *first line of defence*, scouring the posted material for words or combinations of words previously defined. This tool, while uncomplicated, will generate millions (to say the least) of false positive signals[22]. The identification of words or sentences in itself will surface potentially violative content but finding the actual violation takes more than that. This is where AI solutions, such as Machine Learning come in. Briefly, Machine Learning (ML) is a subcategory of AI in which algorithms are employed to "*automatically learn insights and recognize patterns from data*"[23] and

[18] Even if the material is not illegal, many companies chose to disallow it on their platforms. For example, Facebook or Instagram (both belonging to Meta) do not allowed pornographic material of the platform but Twitter does.
[19] In platforms like TikTok, Twitter or Facebook, "first line" moderators should be able to examine and act in less than 2 minutes.
[20] Unpublished, privatised or removed.
[21] Content moderation teams frequently feature policy area-specific moderators (minor safety, terrorism, and the like)
[22] Only words are flagged, sentiment is not detected.
[23] Columbia Engineering

"*imitate the way that humans learn, gradually improving its accuracy*"[24]. Google, which is, potentially, the company with the largest datasets of information, has created the tool Vision AI which identifies objects, people, or animals within an image. For example, in 2014, Youtube engineers developed CSAI (Child Sexual Abuse Imagery) Match which is a CSAM-specific tool for video content. Essentially, it looks for content matching the video hash in the dataset. Hashing "*is the process of taking a big volume of data and reducing it into a small volume of data by assigning a unique numerical identifier to a file, a group of files, or a portion of a file*" (Thorn, 2016). Succinctly, hashing allows for the alphanumeric fingerprinting of a specific image[25], even if altered (Thorn, 2016) enabling the quick, nearly unequivocal, identification of illegal content. This is how Thorn, Youtube (CSAI Match), Twitter (Microsoft's PhotoDNA[26]) or Apple's NeuralHash[27] *"find and remove* (known) *content quickly*" (Thorn, 2016). These tools and hashes are also shared with non-governmental organizations such as the US's NCMEC (National Center for Missing and Exploited Children) and the UK's IWF (Internet Watch Foundation). In fact, in 2019, both NGOs made a data-sharing agreement, essentially combining their hashes to better identify child abuse images/videos and fight the sharing of illegal content.

But what of unknown or new CSAM? Google's Vision AI can also identify children and the model may have learned to flag images with children in states of undress, or the content may have been reported by a user or the very victim. Despite the increased accuracy of the tools, at the end of the long line of AI evaluation, there must be a person to make a judgment call. Once that call is made and be it the case that the image or video is identified as CSAM and reported, the content can be hashed and added to the dataset. Another concerning trend is the digital manipulation of images to create *deepfakes*[28]. While face-swapping apps began as an entertaining gimmick, it didn't take long for users to employ it for far more nefarious ends. From the propagation of misinformation[29] to celebrity or harassment-aiming pornography[30]. Nonsurprisingly, this trend also extended to child abuse material, while fortunatily it remains fairly uncommon (WeProtect, 2021).

---

[24] IBM

[25] Not to be confused with Content-Based Image Retrieval (such as reverse image search) which is "*a set of techniques for retrieving semantically-relevant images from an image database based on automatically-derived image features* […] *extracting features of every image for its pixel values*" (Alkhazraj, 2017)

[26] https://www.microsoft.com/en-us/photodna

[27] https://www.apple.com/child-safety/

[28] Cambridge's online dictionary defines deepfakes as a video or sound recording that replaces someone's face or voice with that of someone else, in a way that appears real.

[29] For example, in 2020, a Belgium branch of a worldwide environmental movement, shared a deepfake of Prime Minister Sophie Wilmès promoting the idea of a link between deforestation and COVID-19 (Holubowicz, 2020).

[30] Ellis, 2018

With respect to the detection of online grooming different tools and strategies must be implemented. As identified by Thorn, and because of US' Children's Online Private Protection Rule (COPPA)[31] which prohibits the capture of information by internet platforms of minors under the age of 13 without parental permission, children, more often than not, particularly between the ages of 9 to 12 years old (61%) admit to pretending to be older online so they can register and access a social media platform or website. On the other hand, offenders may also pretend to be younger when engaging with a child (de Setisteban et al's 2018). Hence, identifying who's who in this context becomes particularly challenging, although not impossible.

Research and intelligence have indicated (WeProtect, 2021; [Name withheld], 2022) that online grooming tends to take place in private chat conversations, where the offender is unrestrained from attentive eyes and public call outs. In an effort to identify offending behaviours and protect minors from online grooming (among other illegal activities), countries have begun to rethink the demands made from social media platforms. In 2021, India has introduced Information Technology Rules which instructs intermediaries (social media and OTT streaming platforms[32]) to record user information, "*erase contentious content fast and assist investigations"* (Reuters, 2021) by preserving "*information for a period of one hundred and eighty days after any cancellation or withdrawal of his registration*" (Rule 3(1)(h)) and "*in the nature of messaging shall enable the identification of the first originator of the information on its computer resource*" (Rule 4(2)), despite the existence of a case for increased "*political control*" (Gupta, 2021). The rule also requires these intermediaries to "*deploy technology-based measures, including automated tools or other mechanisms to proactively identify information that depicts any act or simulation in any form depicting rape, child sexual abuse or conduct, whether explicit or implicit*" (Rule 4 (4)).

The UK has also drafted and brought to Parliament a bill in 2022 (Online Safety Bill) which "*requires technology companies to protect their users from illegal content such as child-abuse images*" (BBC, 2022) by removing it as well as instructing platforms to age verify users and employ tools and strategies to prevent children from being exposed to harmful content (UK Government, 2022).

Similarly, in May 2022, the European Commission has proposed Rules to Prevent and Combat Child Sexual Abuse. The proposal outlines that platforms must perform mandatory risk assessments and mitigation efforts, essentially ordering these companies to self-moderate (detect, report and remove material) in accordance with EU's law as well as

---

[31] 15 U.S.C. § 6501 Online Version, https://www.govinfo.gov/content/pkg/PLAW-105publ277/html/PLAW-105publ277.htm (Last visited Oct 15th, 2022)

[32] e.g. Netflix

employ strategies to reduce expose to online grooming by ensuring "*that children cannot download apps that may expose them to a high risk of solicitation of children*" (European Commission, 2022)[33].

WeProtect's 2021 Global Threat Assessment has reported that, while the majority of internet companies that answered the survey, use hashing to detect CSAM, only 37% employ similar strategies to tackle OG (WeProtect, 2021). This is particularly concerning, if we consider that avoiding detection in this case is arguably simple. Offenders' use of end-to-end encrypted (E2EE) messaging apps[34], VPNs[35], *burner accounts*[36] or the mere use of more than one platform[37] allows for basic but effective and guileful concealment of their activities from the platforms and authorities. Furthermore, hash-matching techniques or grooming detection algorithms "*are not readily deployable within E2EE environments*" (WeProtect, 2021). In 2020, NCMEC reported an increase of nearly 98% in what the NGO defines as online enticement, an umbrella term that encompasses OG and similar behaviours with the aim of committing a sexual offence or abduction(NCMEC data in WeProtect, 2021). That said, the industry is attempting to find and employ solutions to detect and even prevent OG.

Several authors have explored AI's ability to detect online grooming while distinguishing it from a cybersexual talk between adults or teenagers since the isolated application of a BoW would return high amounts of false positives. Powell et al (2021) were able to identify "moves" employed by offenders in chatlogs and proposed the development of scanning tools able to establish associations between explicit sexualized speech and sentences to overcome the victim's resistance, rather than the use of a BoW (Bag of Words) to identify potential grooming activity. In 2016, Cardei and Rebedea proposed a two-stage classification system to identify potential predator behaviour online, where the first stage detects suspicious vocabulary and eliminates innocuous conversations, the second attempts to pinpoint inappropriate sexual conduct, through the use of the BoW model, behavioural features that quantify the activity of a given user and interactional features to distinguish between the predator and the victim. Kinzel (2021)'s linguistic approach found distinguishing

---

[33] The proposal has raised some concerns as it may undermine the safety and privacy of end-to-end encryption (E2EE) by posing an obligation to scan private communications (Burgess, 2022).

[34] e.g. What'sApp or Telegram, whose code allows for "*encrypted data is only viewable by those with decryption keys*" (IBM), the sender and the receiver.

[35] Virtual Private Networks that encrypt the connection and "jump" between IPs often across continents. As such, identifying the actual location of a user becomes virtually impossible.

[36] Social media accounts generally created to post or engage with others relatively anonymously and refrain from linkage to the user's "official" account.

[37] NCMEC has found that, in an analysis of received reports for online enticement, 42% of the offenders used more than one platform when attempting to engage minors (NCMEC data in WeProtect, 2021).

features of grooming language and was also able to draw recognizable associations between keywords used and the intentions of the writer. In light of this research, the author proposed a set of 116 words[38] that, when found in a unique chatlog, may indicate the presence of grooming. This type of research is fundamental for the construction of datasets to train ML tools. Similarly, Cano et al (2014) analysed the issue from a machine learning perspective and the authors were able to gather significant results regarding AI's grooming identifying capabilities in the detection of different stages of online grooming (trust development, grooming and seeking for physical approach) in online chat discourse. In 2019, Anderson et al proposed a system consisting of a data retrieval tool (BoW) coupled with AI feature selection mechanisms capable of automatic distinction between grooming and non-grooming interactions.

At this time, to the best of my knowledge, there is only one online grooming detection tool: 2020's Microsoft Project Artemis. This technique allows for the detection, mitigation and reporting of online grooming occurrences. The tool was developed to tackle this type of abuse on Microsoft's Xbox chat feature (*The Verge*, 2020) and was later shared with the industry to increase its precision and reliability. The AI tool was trained via a dataset of confirmed online grooming instances, spots, in real time, specific words, statements and speech patterns which are flagged and rated for risk and then sent for human moderators to review (Harris, 2020). If the moderator confirms abusive conduct is taking place, the offending user is then reported to NCMEC (Gregorie, 2020). Additionally, because the tool allows for a *tunning* of its features, companies may decide for automated action (Harris, 2020) which (although what that may entail was not clearly stated by those involved) could result in a swift block of the offender's activity on the platform and account removal. Similarly, Swansea University's Project DRAGON-S' project[39], headed by research Nuria Lorenzo-Dus[40], started in 2021 and aims to develop detection (DRAGON-SPOTTER) and education tools (DRAGON-SHIELD). Automated grooming detection is still very much in its infancy and further adoption of tooling and widespread use and enhancement (language and colloquial terms, emojis or keyword obfuscations are ever evolving and the database of known terms needs to be systematically monitored and updated) are needed to construct a reliable instrument.

While these remain in the works, risk intelligence and cybersecurity companies prove to be helpful allies in the detection of abusive conduct. These entities devote their time to investigating and gathering information regarding a variety of topics and concerns, including

---

[38] Such as *sweetie, foreplay, tummy, jail* or *cop.*

[39] https://www.swansea.ac.uk/project-dragon-s/

[40] Chair of Applied Linguistics of the institution and whose work has been been previously mentioned in this thesis

the identification of child predator networks and individuals. Because these offenders may evade authorities with some level of ease (through the use of tools such as VPNs), they often “jump” between platforms with a preference for those with less moderation activity. An added concern is the fact that many of these networks mostly operate in private and small message groups which not infrequently require the sharing of incriminating material as an entry token ([Name Withheld], 2022) as an evasion technique. As such, tracking of so-called *bad actors* may also prove to be an effective method of tackling online offending activity.

Shared responsibility

The online safety of children is a responsibility that must be shared across the board. Policy-makers, tech companies, law enforcement and parents, guardians, teachers and society as a whole. Education is undoubtedly the first needed step for a successful crusade on this matter to help children and those closest to them realise the potential danger faced while online and bridge any communication gaps so that the victims feel safe in coming forward and reporting. Notwithstanding, tech platforms can still play a bigger role in the prevention and mitigation of these occurrences.

In 2019, the Australian Commission for online security established a set of ruling principles that must be applied by service providers, online platforms and “any technological tool, product, device or service that enables interaction within the general population” to ensure the safety of all the (Australian) netizens. To these, the commission gave the name *Safety by Design* (SbD) which, as the name suggests, looks for the implementation of “*practical, realistic and achievable*” measures on the online fabric of the virtual landscape (eSafety Commissioner, 2019). Three high level principles are outlined:

(1) *Service Provider Responsibilities* — presence of accountable teams for the development and updating of policy, guidelines and terms of use that are in line with governing laws as well as their fair and documented implementation and moderation, coupled with the existence of triaging tools, clear and easy-to-access escalation and user reporting paths and other detections methods while also establishing protocols for reporting any legally contemplated material or behaviour to law enforcement or other governing entities and designing all of these processes with user’s privacy and safety in mind;

(2) *User empowerment and autonomy* — provision of features that permit the user to manage their own online safety and engagement but setting the most secure settings by default, the definition and implementation of explicit consequences for violations of the digital agreement made at the time of engagement with the platform while simultaneously proving users a chance to appeal the decision, and evaluation and preemptive built-in mitigation processes for potential risks or harm to users upon the development of a new feature;

(3) *Transparency and accountability* — commitment to the fair and unbiased implementation and treatment of the users must be a core feature of the moderation efforts and that users can easily locate and read through the set guidelines and terms, engagement of users and external consultants in the development and maintenance of up to date and effective safety standards, and publishing annual reports concerning reporting statistics and consequent moderation efforts.

Any mainstream, surface web platform features the above-mentioned principles, albeit to varying degrees. When signing up for a platform, users are prompted with a terms agreement form that must be completed in order for the account to be created, regardless if these are read thoroughly or not. This provides the platforms with firm grounds upon which the community and content can be moderated. The consequences of violation of the established policy are also generally outlined in the platform's terms of use/service.

There are further examples of SbD in the industry, particularly by large platforms. For instance, Youtube Kids[41] (the child oriented version of the mainstream video platform) has a myriad of safety features such as the inability to leave comments on videos[42], automated privatisation of videos created on the accounts or disabling of "overly commercial content" (EndGadget, 2021). Youtube, Instagram or Discord all feature built-in reporting pathways that will give the reporting users opportunity to raise their concerns anonymously to the Trust & Safety teams. Some platforms share further details into their moderation processes to provide users with specific insights regarding the procedures (such as Facebook[43] or Patreon[44]). Other companies provide in-app education, such is the case for Apple's Neural Hash tool which, upon recognizing an image featuring sensitive material, it will prompt users with pop-up educational messages[45] allowing the viewer to refrain from opening potentially egregious content. Similarly, Pornhub, the world's largest pornography website, has launched a dissuasion feature (reThink Chatbot) which warns users when they try to search for illegal material on the platform, prompting with a pop-up chatbox message alerting users that they are search for egregious material (Farrel, 2022) and further provides anonymous chatlines to seek help and deter users from committing offences (IWF, 2022). Path disruption methods may also be relatively simple measures that, at the very least, attempt to dissuade curious, distracted or willing individuals from engaging in virtual, hands-off offences

[41] https://www.youtubekids.com/
[42] This measure was implemented due to the concerning presence of *softcore* peadophilic rings on the platform (Endgaget, 2019)
[43] How Meta enforces its policies
[44] How Patreon moderates content
[45] Child Safety - Apple

(WeProtect, 2021). Furthermore, OSTIA[46] members have partnered with the EU to develop an online age verification system to ensure the safety of children online and seek parental consent while upholding GDPR and country specific laws (OSTIA, 2021), coined euCONSENT[47]. The system aims to verify the user's age without the need to find further identity information. As the tool is still in its development stage, further information concerning the exact procedures are still unknown. Live Streaming is also a popular way to distribute illegal material, be it CSAM or SG-CSAM (WeProtect, 2021) and, due to its transient nature (provided the material is not recorded) requires immediate detection and mitigation. Hashing tools are not suitable to address these concerns. However, with its increasing popularity and consequent abuse, companies have begun to develop initiatives to fight this trend. SafeToNet[48] is a cybersafety company that designed the SafeToWatch tool which detects (through an array of AI features), in real time, inappropriate content and behaviour (CSAM, bullying, grooming or aggression) and prompts the users with in-app messages. When the app detects sexual material, the device's microphone and camera are restricted and can be employed in E2EE software (WeProtect, 2021).

Member companies of the Tech Coalition organisation[49] share periodical transparency reports that present data related to the number of reports received and accounts warned and removed, for example:

- Between April and June 2022 Tik Tok removed over 113 million videos from the platform nearly 5 million of which have been removed by an automated system. The shared data suggests circa 49 million videos were removed for Minor Safety concerns, 75,6% of which were removed for CSAM, 2,4% for sexual exploitation of minors and 1,9% for grooming behaviours. Additionally, the company removed more than 20 million accounts for suspicion of its owners being less than 13 years of age.
- In the first quarter of 2022, Discord banned 826,591 accounts for Child Safety concerns, 718,385 of which for sharing CSAM (and 22,499 servers) and 10,695 have been reported to NCMEC. The platform provides further information about the type of content that led to the removal: 10,641 of those reports were images or videos, many flagged through Microsoft's PhotoDNA, however only 54 accounts were removed for grooming.

---

[46] Online Safety Tech Industry Association (OSTIA) is a UK organisation of cybersecurity and intelligence companies.

[47] https://euconsent.eu/

[48] https://safetonet.com/

[49] Tech Coalition is a global network of tech companies focused on fighting online child sexual exploitation and feature members such as Cloudfare, Meta or Zoom to name only a few.

- For the same time period, Mega[50] removed 117.6 billion files from their servers with the CSAM being by far the largest violation reason (95%). Over 20,000 folders were disabled which were mostly reported by individual users.

While initiatives such as these are major steps toward a cleaner, safer, child-friendly online environment, the protection of children is a responsibility shared by a myriad of entities. WeProtect's 2021 Global Threat Assessment made suggestions to four focus areas: (a) consistent legislation regarding internet regulation including determining the legal responsabilities of service providers while working closely working with companies to understand state of the art technologies and provide rules such as the SbD principles provided by the Australian goverment; (b) sturdier law enforcement procedures and dedicated agencies to tackle internet-faciliated child abuse and sharing of data to allow for the detection and removal of known sex offenders' accounts; (c) robust cooperation between tech companies such as sharing user intelligence to identify and action known bad actors, content and tooling as well as systematic implementation of SbD and similar principles; and (d) societal education which would include providing parents, guardians, teachers, children and communities in general with infromation on how to recognize inappropriate behaviours, identify children's vulnerabilities, report and assist in the fighting of virtually initiated child abuse.

Lastly, the media has also a role to play in the perception of the issue from society at large. Mainstream coverage of CSAM related initiatives and arrests as well as the creation of thoughtful towards the victims but thought-provoking narratives for all to ensure raised awareness.

The virtual is Real

Online risks are "real world" risks. Moving away from the land *vs* virtual paradigm is not only necessary, it's inevitable. A new generation of professionals who does not feel the need for a distinction between online or offline has emerged. A generation that does not participate in this dichotomy at all has already been born. A generation of digital natives. The internet is no longer just in the desk computer at the corner of a room. It's in our pockets, all day, every day. The world before the internet is no more.

The internet changed how humans interact in a fundamental way and its unique communication opportunities lie ingrained in the vast majority of people. It altered the way we understand the world, in its vastness and its smallness. The digital landscape just like any city has popular spots and less known alleys, neighbourhoods and communities, interest

[50] A file sharing, end-to-end encryption tool.

clubs and museums. The non-entirely erroneous notion of anonymity has fueled lingering curiosities and poked in constructed resilience. Despite this, only relatively recently has the internet started to face some kind of systematic, while still primordial, governance, so much so that for many the WWW at the beginning of an URL might as well still stand for the *wild wild west*. Fortunately, average netizens remain buoyant at the surface of the internet ocean but, as the favoured parks and popular areas become insidiously populated, many, like others before them, may feel the need to explore other venues and obscure quarters where surveillance is scarce or non-existent. The most adventurous will persevere and find ways to reach those regions, but we wouldn't permit our children to walk away with a stranger in the most secure of parks, so why should we allow for them to be approached by unknown adults in this environment, from the very safety of their home? There's a pressing need for digital governance maturing, to understand the internet is not a craze or a trend but a long-lasting, defining communication tool that requires discernment and investigation as well as sturdy, dedicated legislation.

References

(2020). Online Safety Bill to return as soon as possible. BBC. https://www.bbc.com/news/technology-62908598

(2020). *Online Safety Bill: factsheet*. UK Government. https://www.gov.uk/government/publications/online-safety-bill-supporting-documents/online-safety-bill-factsheet

(2022). *Legislation to prevent and combat child sexual abuse*. European Commission for Migration and Home Affairs. https://home-affairs.ec.europa.eu/whats-new/campaigns/legislation-prevent-and-combat-child-sexual-abuse_en

(2022). *What is end-to-end encryption?* IBM. https://www.ibm.com/topics/end-to-end-encryption

[Name withheld]. (2022) Security & Privacy: Detecting Child Predator Communities on Instant Messaging. Unpublished confidential document.

Alexy, E. M., Burgess, A. W., & Baker, T. (2005). Internet Offenders. In Journal of Interpersonal Violence (Vol. 20, Issue 7, pp. 804–812). SAGE Publications. https://doi.org/10.1177/0886260505276091

Alkhazraj, H. (2017). Study for constant-based image relative: A Review. IET Image Processing. IEEE. [Online Version]

American Psychiatric Association. (2013). Paraphilic disorders. In Diagnostic and statistical manual of mental disorders (5th ed.).

Anderson, P., Zuo, Z., Yang, L., & Qu, Y. (2019). An Intelligent Online Grooming Detection System Using AI Technologies. In 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE. https://doi.org/10.1109/fuzz-ieee.2019.8858973

Babchishin, K. M., Hanson, R. K., & VanZuylen, H. (2014). Online Child Pornography Offenders are Different: A Meta-analysis of the Characteristics of Online and Offline Sex Offenders Against Children. In Archives of Sexual Behavior (Vol. 44, Issue 1, pp. 45–66). Springer Science and Business Media LLC. https://doi.org/10.1007/s10508-014-0270-x

Babchishin, K. M., Hanson, R. K., & VanZuylen, H. (2014). Online Child Pornography Offenders are Different: A Meta-analysis of the Characteristics of Online and Offline Sex Offenders Against Children. In Archives of Sexual Behavior (Vol. 44, Issue 1, pp. 45–66). Springer Science and Business Media LLC. https://doi.org/10.1007/s10508-014-0270-x

Babchishin, K. M., Karl Hanson, R., & Hermann, C. A. (2010). The Characteristics of Online Sex Offenders: A Meta-Analysis. In Sexual Abuse (Vol. 23, Issue 1, pp. 92–123). SAGE Publications. https://doi.org/10.1177/1079063210370708

Bailey, J. M., Bernhard, P. A., & Hsu, K. J. (2016). An Internet study of men sexually attracted to children: Correlates of sexual offending against children. In Journal of Abnormal Psychology (Vol. 125, Issue 7, pp. 989–1000). American Psychological Association (APA). https://doi.org/10.1037/abn0000213

Bailey, J. M., Hsu, K. J., & Bernhard, P. A. (2016). An Internet study of men sexually attracted to children: Sexual attraction patterns. In Journal of Abnormal Psychology (Vol. 125, Issue 7, pp. 976–988). American Psychological Association (APA). https://doi.org/10.1037/abn0000212

Bartels, R. M., & Merdian, H. L. (2016). The implicit theories of child sexual exploitation material users: An initial conceptualization. In Aggression and Violent Behavior (Vol. 26, pp. 16–25). Elsevier BV. https://doi.org/10.1016/j.avb.2015.11.002

Beech, A. R., Elliott, I. A., Birgden, A., & Findlater, D. (2008). The Internet and child sexual offending: A criminological review. Aggression and Violent Behavior, 13(3), 216–228. https://doi.org/10.1016/j.avb.2008.03.007

Bergen, E., Davidson, J., Schulz, A., Schuhmann, P., Johansson, A., Santtila, P., & Jern, P. (2014). The Effects of Using Identity Deception and Suggesting Secrecy on the Outcomes of Adult-Adult and Adult-Child or -Adolescent Online Sexual Interactions. In Victims & Offenders (Vol. 9, Issue 3, pp. 276–298). Informa UK Limited. https://doi.org/10.1080/15564886.2013.873750

Black, P. J., Wollis, M., Woodworth, M., & Hancock, J. T. (2015). A linguistic analysis of grooming strategies of online child sex offenders: Implications for our understanding of predatory sexual behavior in an increasingly computer-mediated world. In Child Abuse & Neglect (Vol. 44, pp. 140–149). Elsevier BV. https://doi.org/10.1016/j.chiabu.2014.12.004

Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. In Journal of Computer-Mediated Communication (Vol. 13, Issue 1, pp. 210–230). Oxford University Press (OUP). https://doi.org/10.1111/j.1083-6101.2007.00393.x

Briggs, P., Simon, W. T., & Simonsen, S. (2010). An Exploratory Study of Internet-Initiated Sexual Offenses and the Chat Room Sex Offender: Has the Internet Enabled a New Typology of Sex Offender? In Sexual Abuse (Vol. 23, Issue 1, pp. 72–91). SAGE Publications. https://doi.org/10.1177/1079063210384275

Burgess, A. W., Carretta, C. M., & Burgess, A. G. (2012). Patterns of federal Internet offenders: A pilot study. In Journal of Forensic Nursing (Vol. 8, Issue 3, pp. 112–121). Ovid Technologies (Wolters Kluwer Health). https://doi.org/10.1111/j.1939-3938.2011.01132.x

Burgess, M. (2022, May 11). The EU Wants Big Tech to Scan Your Private Chats for Child Abuse. *Wired*. https://www.wired.co.uk/article/europe-csam-scanning-law-chat-encryption

Calvete, E., Fernández-González, L., González-Cabrera, J., Machimbarrena, J. M., & Orue, I. (2020). Internet-Risk Classes of Adolescents, Dispositional Mindfulness and Health-Related Quality of Life: A Mediational Model. In Cyberpsychology, Behavior, and Social Networking (Vol. 23, Issue 8, pp. 533–540). Mary Ann Liebert Inc. https://doi.org/10.1089/cyber.2019.0705

Cano, A. E., Fernandez, M., & Alani, H. (2014). Detecting Child Grooming Behaviour Patterns on Social Media. In Lecture Notes in Computer Science (pp. 412–427). Springer International Publishing. https://doi.org/10.1007/978-3-319-13734-6_30

CARDEI, C., & REBEDEA, T. (2017). Detecting sexual predators in chats using behavioral features and imbalanced learning. In Natural Language Engineering (Vol. 23, Issue 4, pp. 589–616). Cambridge University Press (CUP). https://doi.org/10.1017/s1351324916000395

Children's Online Private Protection Rule, 15 U.S.C. § 6501 (1998). https://www.govinfo.gov/content/pkg/PLAW-105publ277/html/PLAW-105publ277.htm

Columbia Engineering. Artificial Intelligence (AI) vs. Machine Learning.https://ai.engineering.columbia.edu/ai-vs-machine-learning/

Cooper, A. (1998). Sexuality and the Internet: Surfing into the new millennium. CyberPsychology & Behavior, 1(2), 187–193. https://doi.org/10.1089/cpb.1998.1.187

Craven, S., Brown, S., & Gilchrist, E. (2006). Sexual grooming of children: Review of literature and theoretical considerations. In Journal of Sexual Aggression (Vol. 12, Issue 3, pp. 287–299). Informa UK Limited. https://doi.org/10.1080/13552600601069414

de Santisteban, P., del Hoyo, J., Alcázar-Córcoles, M. Á., & Gámez-Guadix, M. (2018). Progression, maintenance, and feedback of online child sexual grooming: A qualitative analysis of online predators. In Child Abuse & Neglect (Vol. 80, pp. 203–215). Elsevier BV. https://doi.org/10.1016/j.chiabu.2018.03.026

DeHart, D., Dwyer, G., Seto, M. C., Moran, R., Letourneau, E., & Schwarz-Watts, D. (2016). Internet sexual solicitation of children: a proposed typology of offenders based on their chats, e-mails, and social network posts. In Journal of Sexual Aggression (Vol. 23, Issue 1, pp. 77–89). Informa UK Limited. https://doi.org/10.1080/13552600.2016.1241309

Dent, S. (2019). Google's new safety measures are designed to protect kids on YouTube, Search and more. EndGadget. https://www.engadget.com/google-unveils-major-safety-changes-for-kids-on-search-you-tube-and-more-130051206.html

Discord. (2022). DISCORD TRANSPARENCY REPORT: JANUARY - MARCH 2022. Discord. https://discord.com/blog/discord-transparency-report-q1-2022

Eke, A. W., Seto, M. C. (2015). Risk assessment of online offenders for law enforcement. In Kurt, R., Quayle, E. (Ed.). *Internet child pornography: Understanding and preventing on-line child abuse* (pp.148-168). New York, NY: Routledge, 2012. [Online Version]

Elliott, I. A., & Beech, A. R. (2009). Understanding online child pornography use: Applying sexual offense theory to internet offenders. In Aggression and Violent Behavior (Vol. 14, Issue 3, pp. 180–193). Elsevier BV. https://doi.org/10.1016/j.avb.2009.03.002

Elliott, I. A., Beech, A. R. and Mandeville-Norden, R. (2013) 'The Psychological Profiles of Internet, Contact, and Mixed Internet/Contact Sex Offenders', Sexual Abuse, 25(1), pp. 3–20. https://doi.org/10.1177/1079063212439426

Elliott, I. A., Beech, A. R., & Mandeville-Norden, R. (2012). The Psychological Profiles of Internet, Contact, and Mixed Internet/Contact Sex Offenders. In Sexual Abuse (Vol. 25, Issue 1, pp. 3–20). SAGE Publications. https://doi.org/10.1177/1079063212439426

Ellis, E. G. (2018). People Can Put Your Face on Porn—and the Law Can't Help You. WIRED. https://www.wired.com/story/face-swap-porn-legal-limbo/

eSafety Commissioner. 2019. Safety-by-design overview. [Online Version]

European Commission. (2019). How the EU protects your children online. Medium. https://europeancommission.medium.com/how-the-eu-protects-your-children-online-b3f3c3a939fb

European Commission. (2022, May 11). Fighting child sexual abuse: Commission proposes new rules to protect children [Press release]. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_2976

Europol. (2021). Internet Organised Crime Threat Assessment 2021. Retrieved from: https://www.europol.europa.eu/cms/sites/default/files/documents/internet_organised_crime_threat_assessment_iocta_2021.pdf

Eurostat. (2022). Individuals - frequency of internet use. [Data file]. Retrieved from https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=isoc_ci_ifp_fu&lang=en

Farrell, J. (2022). Pornhub is now using AI to persuade people not to search for illegal content. SiliconANGLE. https://siliconangle.com/2022/09/28/pornhub-now-using-ai-persuade-people-not-search-illegal-content/

Geradt, M., Jahnke, S., Heinz, J., & Hoyer, J. (2018). Is Contact with Children Related to Legitimizing Beliefs Toward Sex with Children Among Men with Pedophilia? In Archives of Sexual Behavior (Vol. 47, Issue 2, pp. 375–387). Springer Science and Business Media LLC. https://doi.org/10.1007/s10508-017-1042-1

Greene-Colozzi, E. A., Winters, G. M., Blasko, B., & Jeglic, E. L. (2020). Experiences and Perceptions of Online Sexual Solicitation and Grooming of Minors: A Retrospective Report. In Journal of Child Sexual Abuse (Vol. 29, Issue 7, pp. 836–854). Informa UK Limited. https://doi.org/10.1080/10538712.2020.1801938

Gregoire, C. (2020). Microsoft shares new technique to address online grooming of children for sexual purposes. Microsoft. https://blogs.microsoft.com/on-the-issues/2020/01/09/artemis-online-grooming-detection/

Griffith College Dublin. (2021, November 15). What is Trust and Safety? [Video]. YouTube. https://youtu.be/A65F1_oyPdc

Gupta, A. (2021, February 26). Centre's IT Rules bring answerability in digital ecosystem. But they also increase political control. *The Indian Express*. https://indianexpress.com/article/opinion/columns/it-act-social-media-govt-control-digital-ecosystem-7204972/

Harris, J. H. (2020). Project Artemis: An Overview. Medium. https://medium.com/themeetgroup/project-artemis-an-overview-9ce4174489db

Hirschtritt, M. E., Tucker, D., & Binder, R. L. (2019). Risk Assessment of Online Child Sexual Exploitation Offenders. The journal of the American Academy of Psychiatry and the Law, 47(2), 155–164. https://doi.org/10.29158/JAAPL.003830-19

Holt, K. (2019). YouTube axes hundreds of channels over child exploitation concerns. EndGadget. https://www.engadget.com/2019-02-21-youtube-removes-channels-comments-child-exploitation.html

Holt, T. J., Blevins, K. R., & Burkert, N. (2010). Considering the Pedophile Subculture Online. In Sexual Abuse (Vol. 22, Issue 1, pp. 3–24). SAGE Publications. https://doi.org/10.1177/1079063209344979

Holubowicz, G. (2020). Extinction Rebellion s'empare des deepfakes. Journalism.Design. https://journalism.design/extinction-rebellion-sempare-des-deepfakes/

IBM. What is machine learning? https://www.ibm.com/cloud/learn/machine-learning

ICMEC. (2018). CHILD SEXUAL ABUSE MATERIAL: Model Legislation & Global Review. Retrieved from: https://cdn.icmec.org/wp-content/uploads/2018/12/CSAM-Model-Law-9th-Ed-FINAL-12-3-18-1.pdf

Interpol. (2020). Threats And Trends Child Sexual Exploitation And Abuse. Covid-19 Impact. Retrieved from: https://www.interpol.int/News-and-Events/News/2020/INTERPOL-report-highlights-impact-of-COVID-19-on-child-sexual-abuse

IWF. (2019). Landmark data sharing agreement to help safeguard victims of sexual abuse imagery. https://www.iwf.org.uk/news-media/news/landmark-data-sharing-agreement-to-help-safeguard-victims-of-sexual-abuse-imagery/

IWF. (2022). Internet Watch Foundation, Stop It Now, and Pornhub launch first of its kind chatbot to prevent child sexual abuse. Internet Watch Foundation. https://www.iwf.org.uk/news-media/news/internet-watch-foundation-stop-it-now-and-pornhub-launch-first-of-its-kind-chatbot-to-prevent-child-sexual-abuse/

IWF. (2022). IWF Annual Report 2021. Retrieved from: https://www.iwf.org.uk/about-us/who-we-are/annual-report-2021/

Jones, S. J., Ó Ciardha, C., & Elliott, I. A. (2020). Identifying the Coping Strategies of Nonoffending Pedophilic and Hebephilic Individuals From Their Online Forum Posts. In Sexual Abuse (Vol. 33, Issue 7, pp. 793–815). SAGE Publications. https://doi.org/10.1177/1079063220965953

Kemp, S. (2022). DIGITAL 2022: JULY GLOBAL STATSHOT REPORT. DATAREPORTAL. https://datareportal.com/reports/digital-2022-july-global-statshot

Kinzel, A. (2021). The Language of Online Child Sexual Groomers - A Corpus Assisted Discourse Study of Intentions, Requests and Grooming Duration [PhD Thesis, Swansea University].

Kleijn, M., & Bogaerts, S. (2020). Sexual Offending Pathways and Chat Conversations in an Online Environment. Sexual Abuse. https://doi.org/10.1177/1079063220981061.

Kloess, J. A., Beech, A. R., & Harkins, L. (2014). Online Child Sexual Exploitation. In Trauma, Violence, & Abuse (Vol. 15, Issue 2, pp. 126–139). SAGE Publications. https://doi.org/10.1177/1524838013511543

Kloess, J. A., Seymour-Smith, S., Hamilton-Giachritsis, C. E., Long, M. L., Shipley, D., & Beech, A. R. (2015). A Qualitative Analysis of Offenders' Modus Operandi in Sexually Exploitative Interactions With Children Online. In Sexual Abuse (Vol. 29, Issue 6, pp. 563–591). SAGE Publications. https://doi.org/10.1177/1079063215612442

Krone, T. (2004). A Typology of Online Child Pornography Offending. (Trends & Issues in Crime and Criminal Justice; No. 279). Australian Institute of Criminology. [Online Version]

Lanning, K. V. (2012). Cyber 'pedophiles': A behavioral perspective. In K. Borgeson & K. Kuehnle (Eds.), Serial offenders: Theory and practice (pp. 71-87). Sudbury, MA: Jones & Bartlett Learning, LLC.

Lorenzo-Dus, N., Izura, C., & Pérez-Tattam, R. (2016). Understanding grooming discourse in computer-mediated environments. In Discourse, Context & Media (Vol. 12, pp. 40–50). Elsevier BV. https://doi.org/10.1016/j.dcm.2016.02.004

Lyons, K. (2020). *Microsoft tries to improve child abuse detection by opening its Xbox chat tool to other companies*. The Verge. https://www.theverge.com/2020/1/14/21063491/microsoft-tool-artemis-abuse-chat-xbox

Machimbarrena, J. M., Calvete, E., Fernández-González, L., Álvarez-Bardón, A., Álvarez-Fernández, L., & González-Cabrera, J. (2018). Internet Risks: An Overview of Victimization in Cyberbullying, Cyber Dating Abuse, Sexting, Online Grooming and Problematic Internet Use. In International Journal of Environmental Research and Public Health (Vol. 15, Issue 11, p. 2471). MDPI AG. https://doi.org/10.3390/ijerph15112471

Malesky, L. A., Jr. (2007). Predatory Online Behavior: Modus Operandi of Convicted Sex Offenders in Identifying Potential Victims and Contacting Minors Over the Internet. In Journal of Child Sexual Abuse (Vol. 16, Issue 2, pp. 23–32). Informa UK Limited. https://doi.org/10.1300/j070v16n02_02

Marcum, C. D. (2007). Interpreting the Intentions of Internet Predators: An Examination of Online Predatory Behavior. In Journal of Child Sexual Abuse (Vol. 16, Issue 4, pp. 99–114). Informa UK Limited. https://doi.org/10.1300/j070v16n04_06

McCarthy, J. A. (2010). Internet sexual activity: A comparison between contact and non-contact child pornography offenders. In Journal of Sexual Aggression (Vol. 16, Issue 2, pp. 181–195). Informa UK Limited. https://doi.org/10.1080/13552601003760006

McGrath, M. G., & Casey, E. (2002). Forensic psychiatry and the internet: practical perspectives on sexual predators and obsessional harassers in cyberspace. The journal of the American Academy of Psychiatry and the Law, 30(1), 81–94.

McManus, M. A., Long, M. L., Alison, L., & Almond, L. (2014). Factors associated with contact child sexual abuse in a sample of indecent image offenders. In Journal of Sexual Aggression (Vol. 21, Issue 3, pp. 368–384). Informa UK Limited. https://doi.org/10.1080/13552600.2014.927009

Mega. (2022). MEGA Transparency Report — March, 2022 Report. Mega. https://transparency.mega.io/#downloadcurorprev

Merdian, H. L., Curtis, C., Thakker, J., Wilson, N., & Boer, D. P. (2011). The three dimensions of online child pornography offending. In Journal of Sexual Aggression (Vol. 19, Issue 1, pp. 121–132). Informa UK Limited. https://doi.org/10.1080/13552600.2011.611898

Mitchell, K. J., Finkelhor, D., & Wolak, J. (2005). The Internet and Family and Acquaintance Sexual Abuse. In Child Maltreatment (Vol. 10, Issue 1, pp. 49–60). SAGE Publications. https://doi.org/10.1177/1077559504271917

Mitchell, R. C., Bravo, A. P., & Galupo, M. P. (2017). Sexual Desire Among an Online Sample of Men Sexually Attracted to Children. In Journal of Child Sexual Abuse (Vol. 26, Issue 6, pp. 643–656). Informa UK Limited. https://doi.org/10.1080/10538712.2017.1328476

Munro, E. (2011). The protection of children online: A brief scoping review to identify vulnerable groups. [Online Version]

O'Connell, R. (2003). A typology of cybersexploitation and online grooming practices. Cyberspace Research Unit, University of Central Lancashire. [Online Version]

Obar, J. A., & Wildman, S. S. (2015). Social Media Definition and the Governance Challenge: An Introduction to the Special Issue. In SSRN Electronic Journal. Elsevier BV. https://doi.org/10.2139/ssrn.2647377

Ofcom. (2021). Children and parents: media use and attitudes report 2020/21. Ofcom. https://www.ofcom.org.uk/__data/assets/pdf_file/0025/217825/children-and-parents-media-use-and-attitudes-report-2020-21.pdf

OSTIA. (2021). OSTIA members partner with EU to launch cross-border online age verification system in 2022. OSTIA. https://ostia.org.uk/2021/05/19/ostia-members-partner-with-eu-to-launch-cross-border-online-age-verification-system-in-2022/

Paquette, S., & Cortoni, F. (2020). Offense-Supportive Cognitions Expressed by Men Who Use Internet to Sexually Exploit Children: A Thematic Analysis. In International Journal of Offender Therapy and Comparative Criminology (Vol. 66, Issues 6–7, pp. 647–669). SAGE Publications. https://doi.org/10.1177/0306624x20905757

Powell M, Casey S & Rouse J 2021. Online child sexual offenders' language use in real-time chats. *Trends & issues in crime and criminal justice* no. 643. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti78481

Prat, S., & Jonas, C. (2012). Psychopathological characteristics of child pornographers and their victims: a literature review. In Medicine, Science and the Law (Vol. 53, Issue 1, pp. 6–11). SAGE Publications. https://doi.org/10.1258/msl.2012.011133

Prat, S., Bertsch, I., Praud, N., Huynh, A.-C., & Courtois, R. (2020). Child pornography: Characteristics of its depiction and use. In Medico-Legal Journal (Vol. 88, Issue 3, pp. 139–143). SAGE Publications. https://doi.org/10.1177/0025817219898151

Prensky, M. (2001). Digital Natives, Digital Immigrants Part 1. In On the Horizon (Vol. 9, Issue 5, pp. 1–6). Emerald. https://doi.org/10.1108/10748120110424816

Quayle, E. (2008). The COPINE Project. Irish Probation Journal, 5, 65-83. http://tinyurl.com/pwtoe88

Quayle, E., & Taylor, M. (2002). child pornography and the internet: perpetuating a cycle of abuse. In Deviant Behavior (Vol. 23, Issue 4, pp. 331–361). Informa UK Limited. https://doi.org/10.1080/01639620290086413

Rimer, J. R. (2019). “In the street they’re real, in a picture they’re not”: Constructions of children and childhood among users of online child sexual exploitation material. In Child Abuse & Neglect (Vol. 90, pp. 160–173). Elsevier BV. https://doi.org/10.1016/j.chiabu.2018.12.008

Risse, M. (2021). The Fourth Generation of Human Rights: Epistemic Rights in Digital Lifeworlds. In Moral Philosophy and Politics (Vol. 8, Issue 2, pp. 351–378). Walter de Gruyter GmbH. https://doi.org/10.1515/mopp-2020-0039

Schoeps, K., Peris-Hernández, M., Garaigordobil, M., & Montoya-Castilla, I. (2020). Risk factors for being a victim of online grooming in adolescents. Psicothema, 32.1, 15–23. https://doi.org/10.7334/psicothema2019.179

Seto, M. C. (2012). Is Pedophilia a Sexual Orientation? In Archives of Sexual Behavior (Vol. 41, Issue 1, pp. 231–236). Springer Science and Business Media LLC. https://doi.org/10.1007/s10508-011-9882-6

Seto, M. C. (2017). The Motivation-Facilitation Model of Sexual Offending. In Sexual Abuse (Vol. 31, Issue 1, pp. 3–24). SAGE Publications. https://doi.org/10.1177/1079063217720919

Seto, M. C., & Eke, A. W. (2017). Correlates of admitted sexual interest in children among individuals convicted of child pornography offenses. Law and Human Behavior, 41(3), 305–313. https://doi.org/10.1037/lhb0000240

Seto, M. C., & Karl Hanson, R. (2011). Introduction to Special Issue on Internet-Facilitated Sexual Offending. In Sexual Abuse (Vol. 23, Issue 1, pp. 3–6). SAGE Publications. https://doi.org/10.1177/1079063211399295

Seto, M. C., Maric, A., & Barbaree, H. E. (2001). The role of pornography in the etiology of sexual aggression. In Aggression and Violent Behavior (Vol. 6, Issue 1, pp. 35–53). Elsevier BV. https://doi.org/10.1016/s1359-1789(99)00007-5

Seto, M. C., Reeves, L., & Jung, S. (2010). Explanations given by child pornography offenders for their crimes. In Journal of Sexual Aggression (Vol. 16, Issue 2, pp. 169–180). Informa UK Limited. https://doi.org/10.1080/13552600903572396

Sexual Exploitation And Other Abuse Of Children, 18 U.S.C. § 2256. https://www.govinfo.gov/content/pkg/USCODE-2015-title18/html/USCODE-2015-title18-partI-chap110.htm

Seymour-Smith, S., & Kloess, J. A. (2021). A discursive analysis of compliance, resistance and escalation to threats in sexually exploitative interactions between offenders and male children. In British Journal of Social Psychology. Wiley. https://doi.org/10.1111/bjso.12437

Soldino, V., Carbonell-Vayá, E. J., & Seigfried-Spellar, K. C. (2020). Spanish Validation of the Child Pornography Offender Risk Tool. In Sexual Abuse (Vol. 33, Issue 5, pp. 503–528). SAGE Publications. https://doi.org/10.1177/1079063220928958

Suler, J. (2004). The Online Disinhibition Effect. In CyberPsychology & Behavior (Vol. 7, Issue 3, pp. 321–326). Mary Ann Liebert Inc. https://doi.org/10.1089/1094931041291295

The Gazette of India. Ministry of Electronics and Information Technology. 25 February 2021. Archived (PDF) from the original on 26 February 2021 – via archive.org.

Thorn. (2016). Introduction to Hashing: A Powerful Tool to Detect Child Sex Abuse Imagery Online. https://www.thorn.org/blog/hashing-detect-child-sex-abuse-imagery/

Thorn. (2018). Production and Active Trading of Child Sexual Exploitation Images Depicting Identified Victims. [Digital Version]

Thorn. (2021). Self-Generated Child Sexual Abuse Material: Youth Attitudes and Experiences in 2020. [Digital Version]

Thorn. (2022). Online Grooming: Examining risky encounters amid everyday digital socialization Findings from 2021 qualitative and quantitative research among 9-17-year-olds. https://info.thorn.org/hubfs/Research/2022_Online_Grooming_Report.pdf

Tik Tok (2022). Community Guidelines Enforcement Report April 1, 2022 – June 30, 2022. Tik Tok. https://www.tiktok.com/transparency/en/community-guidelines-enforcement-2022-2/

Tozdan, S., & Briken, P. (2015). The Earlier, the Worse? Age of Onset of Sexual Interest in Children. In The Journal of Sexual Medicine (Vol. 12, Issue 7, pp. 1602–1608). Elsevier BV. https://doi.org/10.1111/jsm.12927

Ward, T. (2000). Sexual offenders' cognitive distortions as implicit theories. In Aggression and Violent Behavior (Vol. 5, Issue 5, pp. 491–507). Elsevier BV. https://doi.org/10.1016/s1359-1789(98)00036-6

Ward, T., Polaschek, D. L. L., & Beech, A. R. (Eds.). (2005). Theories of Sexual Offending. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470713648

Webster, S., Davidson, J., Bifulco, A., Gottschalk, P., Caretti, V., Pham, T., Grove-Hills, J., Turley, C., Tompkins, C., Ciulla, S., Milazzo, V., & Craparo, G. (2012). European online grooming project (Final report). European Commission Safer Internet Plus Programme, Tech. Report. [Online Version]

WeProtect (2021). *Global Threat Assessment 2021.* https://www.weprotect.org/global-threat-assessment-21/

Whittle, H., Hamilton-Giachritsis, C., Beech, A., & Collings, G. (2013). A review of online grooming: Characteristics and concerns. In Aggression and Violent Behavior (Vol. 18, Issue 1, pp. 62–70). Elsevier BV. https://doi.org/10.1016/j.avb.2012.09.003

Wild, T. S. N., Fromberger, P., Jordan, K., Müller, I., & Müller, J. L. (2019). Web-Based Health Services in Forensic Psychiatry: A Review of the Use of the Internet in the Treatment of Child Sexual Abusers and Child Sexual Exploitation Material Offenders. In Frontiers in Psychiatry (Vol. 9). Frontiers Media SA. https://doi.org/10.3389/fpsyt.2018.00763

Williams, R., Elliott, I. A., & Beech, A. R. (2013). Identifying Sexual Grooming Themes Used by Internet Sex Offenders. In Deviant Behavior (Vol. 34, Issue 2, pp. 135–152). Informa UK Limited. https://doi.org/10.1080/01639625.2012.707550

Winters, G. M., Kaylor, L. E., & Jeglic, E. L. (2017). Sexual offenders contacting children online: an examination of transcripts of sexual grooming. In Journal of Sexual Aggression (Vol. 23, Issue 1, pp. 62–76). Informa UK Limited. https://doi.org/10.1080/13552600.2016.1271146

Woodroofe, J. (2020). A Fourth Generation Of Human Rights? The Organization for World Peace. https://theowp.org/a-fourth-generation-of-human-rights/

Youtube. Protect your content and online community from child exploitation videos. https://www.youtube.com/csai-match/

Kalra, A., & Phartiyal, S. (2021, February 24). India plans new social media controls after Twitter face-off. *Reuters*. https://www.reuters.com/article/us-india-tech-regulation-idINKBN2AO201