RAYUELA
a fun way to fight cybercrime

Deliverable Report

D2.5 Open Report on Technological Threats Associated to the Cybercrimes Considered

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 882828. The sole responsibility for the content of this document lies with the author and in no way reflects the views of the European Union.

Document Contributors-

Deliverable No.	D2.5	Work Package No.	WP2	Task/s No.	2.3
Work Package Title		Technological Assessment and IT Threat Landscape			
Linked Task/s Title		IT threat landscape: identification of most common online threats for children and young adults			
Status		Final	(Draft/Draft Final/Final)		
Dissemination level		PU	(PU-Public, PP, RE-Restricted, CO-Confidential)		
Due date deliverable		31/03/2022	Submission date		31/03/2022
Deliverable version		1.0			
Deliverable responsible		UNIVERSIDAD PONTIFICIA COMILLAS			

Contributors	Organization	Reviewers	Organization
Jaime Pérez	COMILLAS	Pedro Vicente	PJ
Gregorio López	COMILLAS	Jelle Janssens	UGent
Mario Castro	COMILLAS	Violeta Vázquez	Zabala
Rafael Palacios	COMILLAS		
Sonia Solera	UPM		
Mario Vega	UPM		
Manuel Álvarez-Campana	UPM		
Tobias Jacobs	NEC		
Julia Gastinger	NEC		

INDEX

List of Abbreviations

Abbreviation	Description
AI	Artificial Intelligence
API	Application Programming Interface
LEA	Law Enforcement Agencies
LGBTQ	Lesbian, Gay, Bisexual, Transgender, and Queer
ML	Machine Learning
NLP	Natural Language Processing
P2P	Peer to Peer
SVM	Support Vector Machines
RAYUELA	empoweRing and educAting YoUng pEople for the Internet by pLAying

Disclaimer

This document is part of an internal deliverable of the RAYUELA project, funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 882828.
The sole responsibility for the content of this document lies with the authors and in no way reflects the views of the European Union. No personal comments or opinions of the researchers are included in this document, only the results obtained from the research carried out. In no way has any kind of benefit been obtained.

1. Executive Summary

Task 3.2 “IT threat landscape: identification of most common online threats for children and young adults” represents the main interface and contact point between WP2 and WP1 and this deliverable is one of the main outputs of the interdisciplinary work carried out between WP2 and WP1. In WP1, the main human factors affecting the cybercrimes considered in the RAYUELA project (i.e., cyberbullying, online grooming, human trafficking, and misinformation) are analyzed from different perspectives (e.g., psychological and sociological perspectives). In contrast, the main goal of this WP2’s deliverable is to analyze the role and impact of technology and technological threats on such cybercrimes. The methodology followed to achieve this objective combines conducting a literature review with collecting first-hand information through surveys. In particular, we have conducted two surveys to gather insightful information from two main sources: (1) the experts participating in the RAYUELA project working in Law Enforcement Agencies (LEAs) and educational institutions; (2) adolescents, to better understand their Internet habits, the applications and devices they use, and the potential relationships with the considered cybercrimes. The latter survey has been also used in WP1 to better understand the human factors affecting such cybercrimes. This analysis has been included in D1.7 “Open report on Victim and offender profile description report”. It is worthwhile to highlight that, while carrying out the research in WP1 and WP2, it was identified the necessity of conducting such a survey, which complements and goes beyond the outcomes that the consortium committed to deliver as part of WP1 and WP2 of the Grant Agreement.

The report is organized as follows.

Section 2 introduces the context and motivation for the report within the RAYUELA framework, together with the objectives and a summary of the content.

In Section 3, we analysed the main cybercrimes affecting minors: cyberbullying, online grooming, human trafficking, and misinformation, as well as the signs that may indicate their occurrence and the possible adverse effects on the victim. This section is based on the research work carried out in WP1 and aims to provide the context required to understand the rest of the deliverable, making it self-contained.

Section 4 provides a quantitative picture of how European children and adolescents use the Internet and how the COVID-19 pandemic has affected these habits based on recent reports. In addition, we conducted an in-depth analysis of the role of technology (e.g., social networks, messaging applications) in the considered cybercrimes, embedded in an analysis of the modus operandi where appropriate. We also review proposals found in the literature to detect the cybercrimes using technological solutions and briefly discuss their usefulness in real-world cases.

In Section 5, we conducted a survey distributed among RAYUELA consortium members with first-hand experience in these cybercrimes (i.e., LEAs and educational organizations). This section shows a summary and visualizations of their responses, what we consider relevant to this report and an opportunity that only multidisciplinary projects like RAYUELA can have.

In Section 6, we conducted a survey distributed among adolescents to better understand their Internet habits, the applications and devices they use, and the potential relationships with the considered cybercrimes. This represent an addition to previous version of this deliverable (D2.4). This section shows a summary and visualizations of their responses focusing especially on the technology-related questions. As it has already been mentioned, the responses related to human factors affecting the considered cybercrimes are included in D1.7. Only responses from Spain are shown, since it is the only country where the distribution of the survey has been completed so far. We would like to gather responses from other European countries, which will provide useful insights for the game development and for the analysis of the gathered data, as well as for generating scientific publications.

In Section 7, we discussed the role that technology and AI may play in improving the online experience of minors. We analyse their limitations and most relevant challenges found in the reviewed literature and provide our opinions as technology experts. This section also summarises the main takeaways that we consider pertinent for the design of the RAYUELA serious game. Finally, Section 8 draws the main conclusions of the work carried out in the context of the deliverable.

The central insight we can draw from this deliverable is establishing which technologies and social networks are used (and how) in the cybercrimes under consideration. In cases of cyberbullying and misinformation, these are committed entirely on social networks and messaging applications. In online grooming, criminals target and contact the victim through social networks and gaming chats and then move the conversation to private chats and encrypted messaging applications. Human trafficking is more heterogeneous than the others, although similar to online grooming. The dissemination of sensitive material obtained by criminals is usually done through P2P networks and darknet forums. This analysis will have significant relevance when building the adventures of the RAYUELA game and will allow us to develop them coherently with real-world cases.

We also consider highly relevant for the project the data shown at the beginning of Section 4 (on the online habits of European minors), in Section 5 (where we collect valuable information from RAYUELA's expert members regarding the use of technology in the cybercrimes considered), and in Section 6 (where we collect first-hand information from adolescents regarding their online habits and used of apps and devices).

Regarding the role of technology in improving the online experience of minors, we are convinced that AI will be key. However, some significant issues must be addressed to be a helpful tool in real-

world cases, being ethical aspects of capital importance in this sense. This is discussed extensively in Section 7. Finally, Section 8 draws the main conclusions from the work carried out.

2. Introduction

Society's progress towards digitalisation has brought with it many advantages. However, it has also helped criminals find new ways to commit their crimes. Moreover, the ubiquity and easy access to digital platforms make it easier for criminals to access minors to perpetrate the offences. For this reason, in this report, we will conduct an exhaustive analysis of how criminals use technology in the cybercrimes considered in the RAYUELA project.

The methodology used for this purpose includes both carrying out a systematic literature review and gathering first-hand information though two surveys. The first one gathers informationfrom the experts participating in the project working in LEAs and education institutions from different areas in Europe. The second one gathers information from adolescents to better understand their Internet habits, the applications and devices they use, and the potential relationships with the considered cybercrimes.

The main content covered in this report is summarized as follows. We review the available literature and draw general conclusions about the role of technology and technological threats in such cybercrimes. We also examine whether the dangers encountered relate to the victim's age, gender, or other sociological factors to try to cross-validate our findings with WP1, where the issue is addressed in detail. We include first-hand information gathered from members of the RAYUELA consortium who work with children or in LEAs. In addition, we gather information from minors though a survey conducted in schools, asking them about their online habits and experiences. All this information will be necessary for creating the stories of the RAYUELA serious game with veracity and coherence with the real world. Finally, we discuss (from our position as technology experts) the role that technology, particularly Artificial Intelligence (AI), can play in improving children's online experience.

This report is one of the main outputs of the interdisciplinary work between WP1 and WP2 from the RAYUELA project.

- While WP1 analyses the main human factors affecting the cybercrimes considered in the project, WP2 supports this work by analyzing the role and impact of online technologies, such as social networks or messaging applications, in such cybercrimes. This contribution from Task 2.3 is included in the deliverable D2.4, and its open version D2.5.
- While WP2 analyses the vulnerabilities and threats associated with connected devices widely used by minors from a technological perspective, WP1 supports this work by identifying the main human factors that may affect the impact of such technological cybersecurity risks and threats. This contribution from Task 2.3 is included in deliverable D2.2 and its open version D2.3 to complement the technological analysis and tests.

3. Cybercrimes Analysis

The following section aims to provide a brief but comprehensive framework concerning the cybercrimes under consideration. This context is based primarily on the research conducted by WP1, which will also be openly available (open deliverable 1.7) and is compiled here for the sake of completeness and to provide the required context to better understand the rest of the report.

3.1. Cyberbullying

Cyberbullying is "an aggressive, intentional act carried out by a group or individual using electronic forms of contact, repeatedly or over time against a victim that cannot easily defend him or herself" [14]. Although this behaviour extends traditional harassment, cyberbullying is an intentional act carried out through the facilities offered by new technologies and catalysed by social media where the attacker can easily hide under a false identity. To carry it out, the attacker does not modify the modus operandi but the medium, which complicates its early detection since a victim can suffer cyberbullying at any time in any place.

According to data from the ANAR Foundation [33], the risk of cyberbullying victimization on minors peaks between the ages of 13 and 14. However, **age** is not a limiting factor; cyberbullying still occurs in other age groups, although somewhat to a lesser extent. In the bullying case, there do not appear to be significant differences in the gender of the victims. However, in cyberbullying, it seems that girls suffer it more frequently. In addition, some studies show a significant correlation between the victimization of cyberbullying and online grooming [83].

The **symptoms** that indicate that cyberbullying occurs can be found in the victim and the attacker and are often related to **abrupt changes in behaviour**. On many occasions, these symptoms are often only visible when the cyberbullying is in an advanced phase. Early signs of a **victimised** minor are negative emotions and stress, in some cases leading to depression in advanced phases of cyberbullying. The mental and psychosocial consequences on the victim are multifactorial [58], so while one cannot point to a single cause, individuals with pre-existing mental health problems are more likely to suffer damaging mental health consequences after being victims of cyberbullying. Some of the main consequences of cyberbullying are shame, decreased self-esteem, isolation, depression, or even suicide attempts, significantly intensified in adolescents when they do not know the cyberbully. In the **attacker**'s case, there are not always apparent symptoms of cyberbullying, yet in the vast majority of cases, the bully also maintains a (more or less) close relationship with the victim (e.g., they go to the same class) [33].

3.2. Online Grooming

Child grooming is a term to describe a process with a series of harassment crimes involving an adult contacting a child or adolescent to gain his or her trust gradually and then engage him or her in sexual activity. In order to be considered a crime, there must be a solicitation of the children for sexual purposes. It may involve establishing a friendship and emotional connection with the child (and sometimes with the family), ofteh through the adult pretending to be another young person, drawing the child into discussing intimate matters and gradually exposing the child to sexually explicit materials to reduce their resistance or inhibitions about sex [64].

Online grooming is just the practice of grooming through the Internet. The core of online grooming consists of exploiting trust to change the child's expectations of safe behaviour and exploiting fear and shame to keep the child silent [81]. This phenomenon regularly leads to various illicit businesses such as child trafficking, child prostitution, cybersex, or the production of child pornography. It can occur quickly or over a long time and has potential negative consequences for the children's psychological, physical, and social well-being. Another related term to be differentiated from online grooming is the so-called sexting, which refers to sending ---voluntarily or not, by deception or grooming--- audiovisual sexual material, such as photographs or videos [20]. Some studies highlight that the most common ages of victims are **between 12 and 15** years old [73] [82]. Concerning gender, the victims tend to be predominantly **girls** (between 69% and 84% depending on the country of the study) [75] [82]. It can be challenging to know if a child is a grooming victim: the signs are not always obvious and may be hidden. Some of the most typical **signs** [81] of online grooming victimization are:

- Having abrupt changes in behaviour and mood
- Sexualised behaviour, language, or an understanding of sex not appropriate for his/her age
- Being excessively secretive about how they are spending their time
- Having an older boyfriend or girlfriend
- Having money or new things that they cannot explain
- Underage drinking or drug-taking
- Being annoyed, withdrawn, or distressed more frequently

Grooming can have both short-term and long-term effects. The psychological impact can last a lifetime. Some of the most typical **effects** [81] of online grooming victimization are:

- Anxiety, depression, post-traumatic stress, difficulty coping with stress, and eating disorders
- Self-harm, suicidal thoughts.
- Sexually transmitted infections, pregnancy
- Feelings of shame and guilt
- Drug addictions
- Relationship problems with family, friends, and partners

3.3. Human Trafficking

The Palermo Protocol [38] defines trafficking in children as "the recruitment, transportation, transfer, harbouring or receipt of a child for the purpose of exploitation." Child trafficking, as defined by the Palermo Protocol, includes three key elements. First, there is the action involved in trafficking (i.e., recruitment, transport, transfer, harbouring, or receipt of a child). Second, there is the purpose of trafficking, which is exploitation and abuse. Third, the forms of exploitation used to exploit the child (e.g., forced prostitution, forced labour and services, and slavery). In many countries, each of these individual elements is a criminal offence by itself and is addressed by different pieces of legislation. However, child trafficking is a legal concept that is separate from its individual elements. A 'comprehensive definition' of child trafficking, according to international standards, brings together these different elements and ensures that the complex 'concept' of child trafficking is defined in national legislation. [38]

This definition of trafficking related to children differs from the broad definition of trafficking in human beings in two critical ways: No violence, deception, coercion, or other fraudulent means are required for a child to be considered a victim of trafficking. Second, the concept of exploitation in child trafficking is broader than for the trafficking of adults. The forms of exploitation identified are those specified in the international definition provided by the 'Palermo Protocol': exploitation of the prostitution of others, other forms of sexual exploitation, labour exploitation, slavery, or practices similar to slavery, servitude, and removal of organs. In one-third of the countries, illegal adoption is also included as a form of exploitation in national definitions of trafficking [38]. In this definition, "child" means any person under 18 years of age.

Child trafficking occurs in virtually all countries in Europe. There is no clear-cut distinction between countries of origin and destination in Europe; in more than half of the countries, trafficking routes lead in both directions, into and out of the country. Children are trafficked across borders but also within countries (internal trafficking occurs in every second European country [38]). Trafficking in children has been perceived mainly in connection with sexual exploitation, but the reality is far more complex. Children in Europe are also trafficked for exploitation through labour, domestic servitude, begging, criminal activities, and other exploitative purposes. [38]

The international definition of child trafficking does not explicitly define exploitation but lists some examples: "Exploitation shall include, at a minimum, the exploitation of the prostitution of others or other forms of sexual exploitation, forced labour or services, slavery or practices similar to slavery, servitude or the removal of organs." [38]

According to data from the ANAR Foundation [82], in the cases of prostitution of minors reported in Spain, the most common **age** range is between 13 and 17, with a lower number of cases involving children under 12. However, the number of reported cases is relatively small. Moreover, according to various studies [82] [30], most human trafficking and prostitution victims are **women**.

Sex trafficking of children is often a hidden crime that happens in plain sight [29]. It is usually tough to identify sexually exploited children. Still, as described by Goodman, Miriam, and Julie Laurence. "Child Trafficking Victims and the State Courts." [29], there is agreement among service providers about general **signs** (or red flags) that offer some guidance in the identification, including:

- Evidence of physical, mental, or emotional abuse
- Inability to speak on one's own behalf
- Inability to speak to an official alone
- Excess amounts of cash on hand
- Tattoos
- Working for long hours with little or no pay
- Presence of an older male or boyfriend who seems controlling
- Loyalty and positive feelings towards a trafficker
- Exhibition of fear, tension, shame, humiliation, and nervousness; and
- Lack of ability or unwillingness to identify as a victim.

Further, Goodman, Miriam, and Julie Laurence. "Child Trafficking Victims and the State Courts." [29] describes some of the elements that a juvenile court judge may see in a case file for a delinquency or dependency case that may indicate that the juvenile is at risk of being trafficked. The indicators all involve a disruptive home or school environment that leaves the juvenile vulnerable with no safe or stable place to go [29]:

- Absence of supervision at home
- Parents who are or have been in prison
- A history of involvement with child welfare
- Multiple foster care placements or schools attended
- A family history of domestic violence
- Frequent runaway or truancy
- A history of alcohol or drug abuse
- Behavioural problems at school or being behind in grade level.

Regarding the **effects**, many children suffer profound and sometimes permanent damage. Child trafficking is a severe violation of human rights that threatens children's survival and development. It denies children their fundamental rights, including the right to education, health, and protection from exploitation and abuse. [38]

The most common outcomes of trafficked children include homicide, suicide, drug overdose, and adult prostitution. Some trafficked and abused children become sexual abusers themselves. Trafficked children are significantly more likely to develop mental health problems, abuse substances, engage in prostitution as adults, and commit or be victimised by violent crimes later in life. [31]

3.4. Misinformation

Misinformation refers to the dissemination of incorrect or misleading information. Disinformation is a particular case of misinformation, where false information has been deliberately created to cause harm [5]. The differentiation between these two terms is usually not trivial, and in many cases, it is never clear whether an event belongs to one or the other category because it is tough to identify the creator's intention. Therefore, it is more accurate to use the term misinformation, as it is more comprehensive [21]. Misinformation can spread rapidly through insufficiently protected media channels, causing damage in the real world. Unfortunately, there is no simple solution to this problem. It takes awareness, concentration, and skills, and tools to handle information securely.

The world of misinformation is complex and nuanced. Often, true elements combine with false ones to create confusing and powerful pieces of misinformation. C. Wardle et al. [22] proposed a widely accepted taxonomy of misinformation, which identifies seven **types** of problematic content:

- **Satire or Parody:** No intention to cause harm but has potential to fool.
- **False Connections:** When headlines, visuals, or captions do not support the content
- **Misleading Content:** Misleading use of information to frame an issue or individual
- **False Context:** When genuine content is shared with false contextual information
- **Imposter Content:** When genuine sources are impersonated
- **Manipulated Content:** When genuine information or imagery is manipulated to deceive
- **Fabricated Content:** New content entirely false, designed to deceive and do harm.

Hate speech is defined as attacking certain people based on their identity, ethnicity, sexuality, gender, or other attributes. Although hate speech is not strictly misinformation, it does base its arguments and ideas on dishonest information and stereotypes.

One of the biggest culprits for the complexity of this problem is the human brain itself due to cognitive biases. Cognitive biases are systematic patterns of deviation in reasoning processes or judgment. These patterns are beneficial in our lives, making it easier to perform any physical or mental activity, even to understand our reality. However, sometimes this ability of the brain leads us to draw erroneous relationships or wrong conclusions. A prevalent example is the “confirmation bias”, which leads people to search for and notice only information that corroborates their position. These biases may be motivated by biological and social reasons. These biases cannot be avoided entirely, yet people can train and learn techniques that facilitate verifying information.

Exposure to misinformation is of much greater concern in children and adolescents. Firstly, because social networks are their primary source of information and entertainment; secondly, their personalities are being formed at that age and are more suggestible, so it is critical to protect them from illicit or dishonest information. If we fail in this task, they will have an unrealistic view of the society in which they live.

4. Role of Technology in the Considered Cybercrimes

In this section, we report on the role played by technology in each of the proposed cybercrimes. We will survey what applications and social networks are most used by minors in Europe and search for their dependency on age, gender, location, and other demographic or socioeconomic factors. In addition, for each cybercrime, we will study the offenders' modus operandi and how they use technology to facilitate their illicit acts.

Regarding the **use of technology and social networks by European minors**, we find sustained growth in the use of the Internet and new technologies among children and adolescents (from 9 to 16 years old). In particular, for the European Union [23], we can draw the following conclusions:

- From 2010, the **time** spent online has almost doubled in many countries.
- **Smartphones** are the preferred access device (i.e., almost anywhere and anytime connectivity).
- The majority of children report **daily use** of the Internet, especially older ones (15-16). Those older ones spend about twice as much time online than those from 9 to 11.
- Their **top activities** are: Watching videos, listening to music, communicate with friends and family, visiting social networks, and playing games online. However, there are considerable differences in the percentages between countries.
- In most countries, more than 50% of all children use **social networks at least weekly**. The **older ones** have a much **greater interest in socialization.** Despite this, many 9 to 11 years old report visiting a social network daily (ranging from 11% in Germany to 45% in Serbia).
- Around **twice as many boys** as girls **play video games online daily**.
- Among adolescents (12-16), the percentage who received sexual messages (**sexting**) in the past year ranged **between 8% in Italy and 39% in Flanders** (depending on the country), and it **increases significantly with age**.
- Regarding **unwanted sexual requests** online, **more girls and older children** experienced this situation.
- The proportion of children reporting **negative online experiences** rises with age. The proportion of children who reported that they told no one about their negative online experiences ranges **between 4% in France and 30% in Estonia**. In most countries, more girls and younger children talk to their parents about their online activities.
- In most countries, about 10% (slightly below) of the children reported being victims of **cyberbullying.**
- The most common harmful content in most countries was **hate messages**, from 4% in Germany to 48% in Poland.
- The **most common** experience related to **data misuse** is getting a virus or spyware. Personal data misuse **increases with age**.

- **Few children** reported they have **gone without eating and sleeping** because of the Internet **daily or weekly**. From 4% in Slovakia to 21% in Flanders, have daily or weekly spent **less time with family, friends or doing schoolwork** because of the Internet.

Which particular applications do minors use and spend their time online? Qustodio, a parental control platform, provides the following data [24] for children between 5 and 14 (Although it should be noted that this study also includes age ranges below the target age range of the project):

- YouTube is the most used application for **video online** (mean of 68 minutes per day), followed by Netflix, Twitch, and Disney Plus.
- Minors dedicate 76% more time in 2020 to **social networks** than in 2019. TikTok already surpasses Facebook and Instagram in popularity (60 minutes per day dedicated more to TikTok than Facebook), followed by Pinterest, Snapchat, and Twitter.
- Roblox, Among Us, Fornite, Brawl Stars, Clash Royale, and Minecraft are the most popular **video games**. However, gaming is highly dependent on the minor's **age**.
- WhatsApp is the most popular **application to communicate**, followed by Skype, Hangouts, Messages, Discord (that has flourished its popularity, a 90% rise compared with 2019), Zoom, and Google Duo.

Furthermore, we cannot ignore the situation that is being experienced due to the Covid-19 pandemic. Recently, the European Union's publications office conducted a report on how minors aged 10 to 18 have experienced online risks during the 2020 lockdown [79]:

- The children reported spending (daily) between **6 hours and 7.5 hours online** during the lockdown (weekdays), although school activities account for more than half of this time.
- **25%** of the children declared experiencing **more routine disruptions** (going without eating or sleeping because of time spent on the Internet) during the lockdown than before.
- **49%** of the children declared having been **cyberbullied** at some point. Among the children who have already been victims of cyberbullying, 44% reported an increase in this phenomenon during the lockdown.
- **75%** of the children reported encountering **misinformation** online. Moreover, **33%** of the surveyed children reported an **increase** in misinformation.
- **60%** of the children reported encountering **hate messages** online, and **33%** of the surveyed children reported an **increase** in this phenomenon during the lockdown.
- Nearly **50%** of the children have reported seeing gory or **violent images** online.
- Nearly **50%** of the children reported **exposure to self-hurting practices**, and **20%** of the surveyed children reported an **increase** in this phenomenon during the lockdown.
- **33%** of the children reported experiencing that somebody **used their personal information in a way they did not like**, and **13%** of all children reported experiencing this phenomenon more during the lockdown.

- Nearly **33%** of the children reported getting a **virus or spyware** at some point. **14%** of all children reported an **increase** in this experience during the lockdown.

The ANAR Foundation, which helps victimized minors in Spain, has also recently published a report [55] on the problems they have detected during the 2020 lockdown:

- **Grooming** cases **increased** by **60.2%** compared to 2019.
- **Child pornography/prostitution** cases **increased** by **41.7%** compared to 2019.
- From 2019 to 2020, the percentage of detected cases where **technology** played a **key role** has increased from 23.2% to **29.9%**.
- The **psychological sequelae** of Covid-19 have **increased dramatically** (e.g., Anxiety by 280.6%, suicidal ideation by 244.1%, low self-esteem by 212.3%, depression/sadness by 87.7%, eating disorders by 826.3%, self-injury by 246.2%)
- **Child poverty** has **increased** by **307.2%**.

4.1. Cyberbullying

Cyberbullying is an intentional and repeated act aimed at causing harm to a vulnerable victim with difficulties to stop or mitigate the offence. The offender uses the technology to his/her advantage, avoiding identification and committing the crime anytime and anywhere.

There are three distinct roles involved in cyberbullying: the **perpetrator** (carries out the action or attack), the **victim** (suffers the attack), and the **bystander**. Bystanders are individuals who are in some way involved in cyberbullying, regardless of whether they act as facilitators of cyberbullying (e.g., by disseminating bullying content), as victim support (e.g., by reporting the bullying), or by standing aside as mere observers. Another critical role outside the three leading roles is the **defender**, who participates in the fight against cyberbullying by preventing, detecting, and mitigating it.

The following section specifies how Information and Communication Technologies (ICT) are used in cyberbullying and the factors that may cause an individual to be involved in cyberbullying (as one of the prominent roles or as a defender in cyberbullying). Finally, it is specified for the use of social media and which technologies have been the most exploited by cyberbullying.

Technology in Cyberbullying

Existing ICT tools and their multiple daily applications raise the impact and scope of traditional bullying. Multiple variables related to personal factors, environmental events, and behavioural patterns increase the possibility of a young person doing or suffering cyberbullying. In particular, the use that young people make of technology, the ease of access, and technological knowledge are the most relevant factors involved in cyberbullying [76].

On top of technology, another determining factor is **personality**. Thus, Ref. [76] identifies the following bullying behaviours in social networks: harassment, denigration, exclusion, deception, and impersonation. In particular, youngsters with higher empathy and coherent moral reasoning tend not to bully but instead tend to side with and support the victim. One aspect that may make it easier for a bystander not to act on the cybercrime they witness is that they know the bully.

The perpetrator takes advantage of technology and often opts for **widely used social media** to conduct the harassment. In general, the Internet and social media offer the anonymity that allows the perpetrator to carry out the act with the security of not being caught. Moreover, these attacks are easily scalable, replicable, and, as mentioned, allow harassment to be carried out at any time and from anywhere. In this sense, bystanders are involved in cyberbullying (either actively or passively) thought the technologies most used at the time, as the perpetrators do.

While technology is at the centre of cyberbullying (negatively), it can also be used to detect, prevent, and mitigate cyberbullying. There are currently multiple systems that combat cyberbullying, such as the one presented in Yao, Mengfan, Charalampos Chelmis, and Daphney-Stavroula Zois. "Cyberbullying Ends Here: Towards Robust Detection of Cyberbullying in Social Media." [8], which studies these two aspects: scalability and frequency of repetition of cyberbullying. In this sense, we differentiate between two cases: if cyberbullying has not yet been committed, it is a matter of **prevention**; if cyberbullying has taken place, it is about the **mitigation** of its effects and, in both cases, about early **detection** of cyberbullying to prevent it from happening or reduce the harm it may cause.

Prevention

On the **prevention** side, initiatives such as Calvo-Morata, Antonio, Manuel Freire-Moran, Ivan Martinez-Ortiz, and Baltasar Fernandez-Manjon. "Applicability of a Cyberbullying Videogame as a Teacher Tool: Comparing Teachers and Educational Sciences Students." [39] propose a serious game to raise awareness among young people by putting them in the role of victims. This type of educational tool not only prevents bullying but also fosters empathy with a real victim. Moreover, according to Hasse, Alexa, Sandra Cortesi, Andres Lombana Bermudez, and Urs Gasser. "Youth and Cyberbullying: Another Look," [1], it is not only necessary to educate in this way, but it is also important to instil a conscious and responsible use of technology. This responsible use is a crucial point that caregivers and parents need to work on from childhood, with specific parenting strategies that work on the controlled use of technology. Often, caregivers and parents are unaware of youngsters' online activities and time spent on Internet, so supervision of such activities may sometimes be necessary. In this way, and with open communication between the parts involved, cyberbullying can be prevented by teaching young people what content is appropriate to share online.

This responsibility does not only lie with caregivers and parents; part of this education is the responsibility of schools. They are also responsible for promoting good use of technology as well as providing an understanding of cyberbullying. To combat it and minimise cyberbullying rates, they have strategies based on comprehensive analyses and student questionnaires on cyberbullying,

cyber-engagement, and, in general, Internet and social media use [59, 43, 23]. However, not only in schools, an exciting area to study is that of video games. Reports such as McInroy, Lauren B., and Faye Mishna. "Cyberbullying on Online Gaming Platforms for Children and Youth." [50] work on the understanding and awareness of cyberbullying in online gaming experiences.

Aizenkot et al. [51] study the relationship between the uncontrolled publication of private information and the likelihood of cyberbullying. This private information often appears on social media profiles, such as real name and surname, gender, age, hobbies, address, etc. The risk of cyberbullying does not end there but involves other associated risks that threaten the individual's privacy. Examples include identity theft and unauthorised disclosure of shared content online.

Social networks and unaccountable online activity fuel all these risks. In recent years, responsible use of social networks has been reinforced with digital safety awareness campaigns from some of the widely known ICT companies, such as Google's 2017 campaign [1]. This campaign includes educational resources for parents and schools and a game for young people aged 8 to 11 that explains how to create a respectful online space interactively. In fact, in 2015, Google began creating educational programs and workshops for young people to understand online safety.

Similarly, Microsoft has been offering online resources on cyberbullying prevention since 2013. In recent years, it has conducted multiple studies and surveys on digital citizenship and analysis of the impact of cyberbullying. This approach is still present in its current policy, and indeed Microsoft has recently joined the global "Power of Zero" campaign to help children and adults learn how to use technology correctly [1].

Mitigation

Cyberbullying can take many forms, one of them is through written content. Thus, there are systems in charge of text processing from Internet users to detect cyberbullying (using, for example, a tweet datasheet [12]). This data is examined word by word, searching for insults or vulgar words, for specific hashtags and emojis, and, as a result, providing qualitative information about the sentiment and emotion of the user. However, in many cases, the tweet context is also considered to obtain more realistic results [42, 47].

Nevertheless, it is also common to find cyberbullying on multimodal platforms so that audio, photos, and video analysis are also carried out to detect cyberbullying [48]. It is also essential to consider other factors such as special events that can generate controversy and lead to discussions, such as sporting events [44]. For example, Facebook removes all content related to cyberbullying by analysing written content, photos, and videos. Their platform has integrated tools that promote user safety, such as removing a user from the friend list, reporting, blocking, or deactivating a user. They even have a social reporting tool to alert Facebook that content may not be appropriate even if it does not explicitly violate the rules. In the area of prevention, in 2017, Facebook launched the "Safety Centre" to provide information on all available resources for young people, educators, and

parents. In addition to the implementation of multiple educational campaigns to combat cyberbullying [1]. Another instance of active monitoring is the Family Safety service deployed by Microsoft that empowers parents and caregivers to monitor their children's Internet use. Like Facebook, the company also enforces anti-bullying policies on its platforms and blocks the accounts of users who violate these rules [1].

Detection

An essential part of cyberbullying **detection** is identifying abnormal behaviour in the perpetrator and the victim; the other important part is identifying the bullying through technology. M. A. Al-Garadi et al. [41] try to combat aggressive behaviour in social media by studying the process of detecting cyberbullying. Specifically, this paper reviews the construction of predictive models of cyberbullying based on Machine Learning (ML). Internet users are continuously exchanging information that generates a large amount of data of varied nature, i.e., raw data. ML algorithms are responsible for finding correlations and complex relationships between these data. In this line of research, Salawu, Semiu, Yulan He, and Joanna Lumsden. "Approaches to Automated Detection of Cyberbullying: A Survey." [40] works on detecting cyberbullying using ML techniques. This reference offers four approaches: (i) Supervised learning, with classifiers such as SVM and Naive Bayes to work with an accurate predictive model; (ii) Lexicon-based approach, with predefined word lists related to the concrete cybercrime; (iii) Rule-based approach, working by matching text to rules; (iv) Mixed-initiative approach. Along the same lines, [7] studies the behaviour of stalkers based on the content they post on Twitter through ML classification algorithms. To do so, they evaluate text, user, and contact networks, concluding that perpetrators post less and, in general, participate less than a regular user. As a result, their accounts tend to be less popular.

In this sense, feature engineering is a crucial factor in ML models. Algorithms learn to discriminate thanks to the input of learning vectors provided to them. These discriminative features, which are used as inputs for the ML classifier, are the basis of the cyberbullying prediction model (especially as the user's age, gender, personality correlate with the risk of suffering cyberbullying). These characteristics are easily extractable from the context of the user's online profile.

Other methods, such as dictionaries containing words associated with cyberbullying [45] (based on word frequencies), help create victim scores. These scores are obtained from the number of hurtful messages sent and received by the user.

Other studies such as Yao, Mengfan, Charalampos Chelmis, and Daphney-Stavroula Zois. "Cyberbullying Ends Here: Towards Robust Detection of Cyberbullying in Social Media." [8] and Chelmis, Charalampos, and Daphney-Stavroula Zois. "Dynamic, Incremental, and Continuous Detection of Cyberbullying in Online Social Media." [25] focus the detection of cyberbullying on the repetition of such messages with abusive behaviour. Their study is based on the number of false cases of cyberbullying that are detected only with a message with aggressive content. The approach offered by the authors is based on timely and accurate detection of cyberbullying in online social

networks. This new approach has two phases. The first involves detecting potential aggressors at the message level, and the second correlating these results with other hurtful messages sent by the potential harasser to determine if it is a case of cyberbullying. This system has been tested on Instagram and Twitter, but there are other similar systems applicable to other social media, such as the one described in Silva, Yasin N., Deborah L. Hall, and Christopher Rich. "BullyBlocker: Toward an Interdisciplinary Approach to Identify Cyberbullying." [6] tested with Facebook but available for most social media, or the one described in Cheng, Lu, Jundong Li, Yasin N. Silva, Deborah L. Hall, and Huan Liu. "XBully: Cyberbullying Detection within a Multi-Modal Context." [9] with Instagram and Vine datasets. Other studies, such as Chelmis, Charalampos, and Mengfan Yao. "Minority Report: Cyberbullying Prediction on Instagram." [10], with a similar approach, conclude that the more comments, the better the detection system will perform, so they work with historical data from Instagram in this case. The report examines the trade-off between the accuracy of results and speed of decision.

Social Media in Cyberbullying

Due to the high penetration of the Internet and social media in the daily lives of young people, the problems associated with the technologies have become much more relevant and worrisome. These technological media offer anonymity and wide availability to the perpetrator and bystander to commit the crime. Despite this, the victim and the aggressor have a direct relationship in many cases [56]. The aggressors use today's most widely used social media to carry out their harassment. The use of social media without proper knowledge or control of the content shared can make users more vulnerable to being victimised. For this reason, as mentioned above, it is essential to promote appropriate use of the Internet and social media, educate caregivers, parents, and children in the respectful and controlled use of the Internet, and generate a safe digital environment. This is mainly due to publishing content with private/sensitive information that can be attacked [51]. Companies like Google, Microsoft, and Facebook have been working on this issue for many years [1], although we cannot claim that they have succeeded in drastically reducing it. These companies have educational resources for young people, caregivers, and parents that fight alongside cyberbullying prevention.

Table 1 shows the most studied social media in the literature regarding cyberbullying. The steep increase in cyberbullying via Twitter explains the high number of studies around it (see Figure 1). This fact may be because Twitter's *design* invites users to discuss topical issues that sometimes generate controversy but, also, there is a bias caused by the data sharing policy implemented by Twitter. On the other hand, Instagram and Facebook have grown significantly in the number of users in recent years. The interaction they offer with users is different from that of Twitter, and in recent years the audience changed the way they use social media, tending to look for those that offer more multimedia resources. Although they can be used to comment on issues, these platforms have a heterogeneous audience that uses them for different purposes. Cyberbullying can be

encountered in many ways other than written content. In fact, in recent years, there has been a growing need to investigate cyberbullying on multimedia platforms such as YouTube. Other social media where cyberbullying has been detected are some known such as Snapchat, Skype, Hello, and MySpace [11, 48].

Research on social media such as TikTok or Twitch can be expected in the coming years due to their increasing use. In this sense, Lowry, Paul Benjamin, Gregory D. Moody, and Sutirtha Chatterjee. "Using IT Design to Prevent Cyberbullying." [52] and Zhou, Yingfan, and Rosta Farzan. "Designing to Stop Live Streaming Cyberbullying: A Case Study of Twitch Live Streaming Platform." [53] studied how security-by-design could mitigate or prevent cyberbullying. Particularly, the work presented in [52] apply a Control Balance Theory (CBT) framework into the IT design process to understand how different design features can affect the prevention of cyberbullying. For its part, [53] is looking at the live streaming platform Twitch to offer design solutions to combat cyberbullying on its platform. The results of [53] showed that there are communication interface designs (streamer-to-consumers and vice versa, i.e., information asymmetries between announcers and viewers) that can promote or inhibit cyberbullying. This type of research focuses on the user's point of view and the development of the technology itself.

Table 1. Evolution of cyberbullying research on each social media

	2017	2018	2019	2020	2021	References
Twitter	✓	✓	✓	✓	✓	[6, 7, 8, 11, 12, 25, 42, 44, 45, 47, 48, 77]
Instagram		✓	✓	✓	✓	[6, 9, 10, 12, 25, 44, 45, 48, 77]
Facebook		✓	✓	✓	✓	[1, 6, 8, 12, 44, 45, 76, 77]
YouTube			✓	✓	✓	[11, 12, 44, 45]
Vine		✓	✓			[9, 48]
Others		✓	✓		✓	[11, 44, 45, 48]

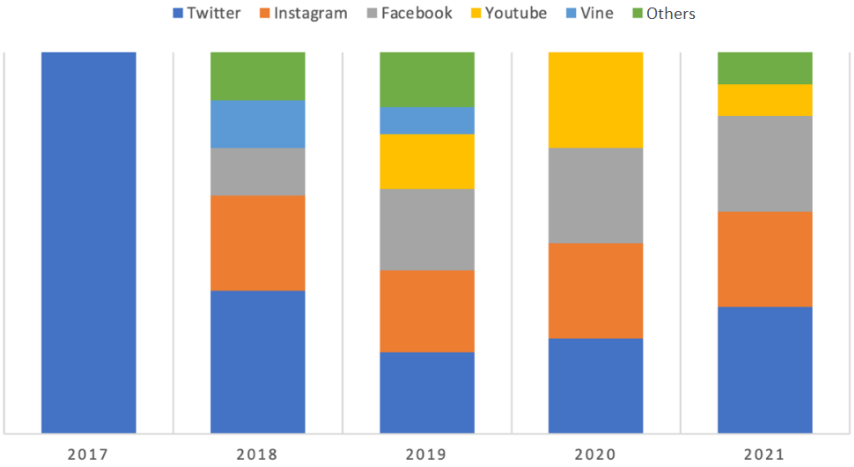

Figure 1. Distribution of reports in research on cyberbullying in social media.

4.2. Online Grooming

Online grooming and child sexual exploitation are constantly evolving phenomena sculpted by technological advances. Online grooming can be committed through any application that allows two or more people to interact with each other (through chats, photos, or videos). Adults seeking to abuse children will go where the minors are, so online grooming can theoretically happen just about anywhere on the Internet. Technology did not create grooming; however, it has increased abusers' reach and opportunity. Today groomers can send a thousand requests in a matter of days and receive 999 declines; it just takes one accepted chat or friend request to open the door.

According to the National Society for the Prevention of Cruelty to Children (NSPCC) [80], Instagram was used in 37% of online grooming cases in the UK, where the platform was recorded. Facebook-owned apps (Instagram, Facebook, WhatsApp) were used in 51% of cases, and Snapchat in 20%. According to the Foundation ANAR, online grooming in Spain grew at a cumulative annual rate of 36.7% between 2013 and 2018 [82].

The **modus operandi** of the criminals can be divided mainly into four phases:

1. Initial Contact / Approach:

   The approach can be either more direct or more indirect. First, the offender identifies potential victims: usually vulnerable children, minors in care or with disabilities, or those at risk of social and economic exclusion [81]. Besides, the offenders usually target regions with higher poverty levels, limited domestic child protection measures, and easy access to children [86]. Contact usually begins in public chat rooms or forums, social networks, game chats, and sometimes through video chat applications (e.g., Chatroulette, Omegle). If the conversation has started in a public chat, the criminal will move the conversation to private chats [61] of any application, but preferably to one that uses end-to-end encryption to avoid surveillance. The offender frequently contacts several potential victims simultaneously, and sometimes each victim by

different means through different social networks, gaming sites, or forums. Therefore, potential victims have a significant risk of having all social networks connected or using the same nickname.

The offender will try to extract all available information from the child's online profiles and environment (interests, usual locations, family conflicts, economic difficulties, problems in their studies, etc.) to find "weaknesses" to attack [69] [63]. Therefore, there is a considerable risk in potential victims' information online, especially if they have public accounts in social networks or accept any person on their friends' list. The offenders frequently lie about their age (pretending to be younger), gender (with the gender to which the minor is attracted), or nationality (to gain the minor's trust further) in their online profiles. Typically, the criminal sends fake photos (of other minors) to the victim posing as them.

The offender will then try to establish a relationship with the minor to get his compliance and "let his guard down". The relationship that the adult will attempt to establish may be romantic, as a friend, as a mentor, or as an authority or dominant figure. Eventually, the adult will begin to introduce sexual topics and sexualise language into the conversation.

2. Sexual Harassment (Escalation and Maintenance):

At this stage, the criminal begins to exploit the vulnerabilities he may have detected in the child: buying gifts or giving money, giving advice or showing fake understanding and attention, attract the victim's interest by offering new experiences, try to isolate the minor from their friends and family, etc. These illegal tactics might induce the minor to establish a dependent relationship with the criminal, which gives the criminal power and control over the minor.

Afterwards, the criminal will try to obtain sexual material from the minor by overcoming any resistance he may offer. Usually, the criminal achieves this through various tactics of psychological abuse or extortion (Manipulation: pressing boundaries, lowering inhibitions, persistence, expressing disappointment or sadness, presenting himself as desperate or in need, begging, pressuring, etc. Harassment: Threatening, bribing, insulting, etc.) [70] [61] [72] [84]. Another common illicit strategy of the offender after grooming (or social engineering) the victim is to coerce and extort them. The offenders use this leverage to gain additional child abuse material, financial gain, physical access to the victim, or extorting the child not to tell anyone about the incident [86].

3. Closure:

Usually, the interactions ended shortly after the online sexual activity occurred. Other interactions continued briefly later, involving several attempts by offenders to reactivate contact or arrange face-to-face meetings [61]. The length of communication between the victim and offender differs depending on whether the offender's goal is to meet the child in person or

engage only in online sexual activity [72]. It is estimated that most grooming cases are not reported to the police due to fear/embarrassment of the minors or their families.

4. Material Dissemination:

This phase occurs in parallel to phases 2 and 3 or even continues after that. It consists of the dissemination of sensitive material obtained by the criminal. Increasing Internet coverage in developing countries and pay-per-use streaming solutions, which provide a high degree of anonymity to the viewer, are fueling the trend of commercial live streaming of child sexual abuse. In 2016, Europol conducted a report on organised crime on the Internet [86] where they analysed, among others, sexual abuse of minors, from which we can highlight the following key findings regarding the technologies used:

- Peer-to-peer (P2P) networks are the most popular platform for the exchange of child sexual exploitation material. However, there are also a growing number of Darknet forums that facilitate the exchange of this material. These forums and the ease of access to Darknet networks (e.g., Tor) lead to an increase in the volume of material exchanged online.
- End-to-end encrypted platforms for sharing media and anonymous payment systems facilitate an escalation in the live streaming of child abuse material.

According to the analysed literature, the Internet (especially social networks) plays a fundamental role in facilitating online grooming. Table 2 shows a summary of the most commonly used technologies in each phase of online grooming. Allowing criminals to obtain crucial information about the minors and attempt to abuse numerous victims simultaneously while hiding their true identity behind the screen. In addition, technology facilitates the lowering of victims' inhibitions, gives the offenders access to minors at anytime and anywhere, and enables extortion with sensitive material through e-mail and social networks (“extortion at scale”) [74]. “The Internet may make some children more vulnerable, but it especially makes vulnerable children more accessible to abusers” [72].

Table 2. Online Grooming - Mapping technology with modus operandi stages.

	Approach	Sexual Harassment	Dissemination
Messaging Apps and private chats	++	++	+
Social Networks	++	++	~
Gaming Chats	++	+	~
Public Chat Rooms or Forums	+	–	–
Video Chat Apps	~	+	~
P2P Networks and Darknet Forums	–	~	++

Legend			
++	+	~	–
Very Frequent	**Frequent**	**Uncommon**	**Never or almost never**

However, technology can also help detect and mitigate cases of online grooming. The world's leading technology companies already have proposals to prevent the use of their platforms for these illicit purposes. Microsoft (together with other companies) developed **Thorn** [2], an anti-grooming tool with two main components:

- PhotoDNA involves creating hash-based digital signatures of the images shared through its platforms and comparing them to hashes from a child pornography database. They can then find copies of such illegal images without the need for complex AI algorithms and computationally efficient.
- A chat scanning tool (text only) assigns a score with the likelihood that the conversation is a grooming case. This tool also does not use complex AI algorithms but relies on more classical techniques for analysing the terms used.

In 2020, Thorn identified daily (on average) 8 victims and 215 files of child sexual abuse. In addition, Thorn informed 155,000 minors daily about risks, prevention, and safety to avoid online grooming.

Facebook [19] and Apple [62] have also implemented tools based on image hashing and text analysis in their chats. Furthermore, they are starting to incorporate AI algorithms to detect multimedia material of sexual abuse of minors and inappropriate interactions with minors. To support this task, Facebook uses the API developed by Google [71] for the same purposes. These initiatives are very positive and have a relevant impact, especially in raising awareness of this problem and advancing the technology to detect it.

Due to the increase in online grooming during the Covid-19 pandemic, the European Parliament has approved temporary rules allowing Internet service providers to use such techniques in their traffic to mitigate this phenomenon [85]. However, many cases of grooming remain undetected. This inconsistency could be mainly because, although in academia grooming detection results are pretty accurate [66] [67] [68] [78], the datasets used to train these models are small, biased (e.g., only in English, from one country, only the conversations that have been released publicly), and were collected a long time ago [65].

Furthermore, detecting grooming cases by text alone is very challenging because initially, the conversations are often polite and almost indistinguishable from a normal conversation. Perhaps the latest advances in Natural Language Processing (NLP), such as OpenAI's GPT-3, could help improve these tools by having a more holistic understanding of language and its meaning. Regarding detecting child sexual abuse multimedia material, algorithms based solely on hashing are easily fooled because a slight change in the material (e.g., crop, zoom, rotate, change some pixels) produces an entirely different hash. To make these hash-based methods more robust, deep neural networks are increasingly being used to distinguish when some material is the same under transformations [54], for example, throughout techniques known as self-supervised learning. However, these deep learning methods are still immature and susceptible to compromise (e.g., adversarial attacks). In any case, hash-based methods can detect only material already in the database used, which significantly limits the scope of the solution.

To take the fight against online grooming seriously, the online platforms through which criminals contact minors (e.g., social networks, gaming chats, messaging apps, public chat rooms) must take the protection of minors more carefully and prioritise it ahead of their commercial interests. In order to advance detection systems, large amounts of data (and metadata) must be collected on detected cases. An in-depth analysis of the patterns used will help to be able to detect cases more effectively. However, close collaboration with police and judicial bodies is needed for this scope, which projects as RAYUELA can promote.

4.3. Human Trafficking

Victims of Human Trafficking

Due to limited data available for Europe, some international data on victims of Human Trafficking is reported in this paragraph. According to some studies on human trafficking in North America [4],

the risk of becoming a victim strongly depends on the social and community context. Some communities in the United States are more vulnerable to becoming victims of child sex trafficking. For example, while Native Americans make up only 11% of New Mexico's population, they account for over 25% of human trafficking victims. African American and Latino minors are also overrepresented when it comes to sex trafficking. Moreover, 50-90% of child sex trafficking victims have been involved in the child welfare system, such as the foster care system. Also, LGBTQ youth face higher rates of discrimination, violence, and economic instability than their non-LGBTQ peers. In particular, homeless LGBTQ youth are three to seven times more likely to engage in survival sex than average homeless persons [4].

In different locales worldwide, the proportion of female to male prostitutes differs to local attitudes, tourism trends, etc. In the US, up to 90% of trafficked children are girls [31]. In Europe, according to data provided by the national authorities, the most significant number of child victims of trafficking was identified in the United Kingdom (2476), followed by Romania (1276) and Bulgaria (196) [35]. Targeted themes include substance abuse, runaway activity, and de-stabilization within the home [30].

There are **two types of underage human trafficking victims**: innocent victims and victims who contribute to their own victimization. Innocent victims are those who mistakenly come across a site used for human trafficking. An example of this would be girls seeking a job such as model, dancer, or nanny lured by human traffickers posing as employment agencies. This also includes a victim using chat rooms or dating sites without being mindful of its particular dangers (the youth and innocence of these underage victims make them more vulnerable).

On the other hand, victims who contribute to their own victimization are the ones who inadvertently do so because of their belief that nothing wrong will happen to them. They may either overestimate their own intelligence or underestimate the trafficker's abilities. Another possibility is that the victim is in a desperate situation and therefore knowingly allows themself to be exploited at the hands of the trafficker with the hope that they can eventually escape when they want to [28].

Research on Human Trafficking

One of the most challenging problems researchers face is the fact that most of the populations relevant to the study of human traffickings, such as victims/survivors of trafficking for sexual exploitation, traffickers, or illegal migrants, are part of a "hidden population" (i.e., it is almost impossible to establish a sampling frame and draw a representative sample of the population) [30]. Even if reliable statistics are not available, it is clear that a significant number of people, predominantly women and children, fall victim to trafficking for sexual exploitation, labour exploitation, or other purposes [34].

For example, according to recent reports in the Lost Kids [114], Lost futures report, the extent of child trafficking in **Germany** appears to be relatively high, whereas authorities in Finland are

unaware of any child trafficking cases. "Any statistics which are released by the authorities must be considered in their proper context", the report asserted, "certain national authorities seem to be unaware of the existence of child trafficking within their borders. This may simply be due to the existence of few cases, or it may be due to the authorities' limited awareness and lack of structure needed to identify and respond adequately to cases. If figures for child trafficking started increasing, it might be that this was not due to an increase in the actual phenomenon but rather an improvement in detection". [34]

Online Platforms used across the phases of Human Trafficking

Victims are often **recruited** from impoverished regions and typically learn about opportunities via word of mouth [37]. One group of traffickers, for instance, used Facebook to browse through user profiles and, based on the information that people shared, selected potential victims who could be courted and tricked into exploitation [38].

In United Nations Office on Drugs and Crime. "Traffickers Use of the Internet; Digital Hunting Fields." [37], three categories of Online platforms are distinguished:

1. **Social Media** – which also includes messengers like WhatsApp or Skype.
2. **Classified Webpages** for advertisement, which refers to generic commercial websites that act as a marketplace for individuals to post advertisements or search for services
3. **Free-standing webpages** which are created by Traffickers themselves

In general, the use of the Internet in child sexual abuse is increasing with greater use of digital media [31]. Online sites for classified ads such as Craigslist (an advertisements website with sections devoted to jobs, housing, for sale, items wanted, services, community service, gigs, résumés, and discussion forums) have already been under intense scrutiny for being used by traffickers [37], as it was found to host many ads for prostitution operations [31].

It has been reported that traffickers that use free-standing web pages are typically able to traffic most **victims**. However, the number of victims per case recorded for trafficking using social media is still significant [38]. Younger victims are reported in cases of trafficking through social media compared to trafficking perpetrated across other platforms [38]. The UOT Human Trafficking Institute study shows exploiters searching sites such as Facebook, Instagram, Twitter, and dating apps like Tinder for posting activity, which might indicate vulnerability [30]. APPG's 2018 report claims that adult services websites such as Vivastreet and Adultwork "represent the most significant enabler of sexual exploitation in the UK" [30].

Approaching and recruitment of victims using digital technologies

According to Goodman, Miriam, and Julie Laurence. "Child Trafficking Victims and the State Courts." [29], recruitment pathways for human trafficking can be categorised into (a) Parents selling children, (b) violence and force, (c) kidnapping, (d) seduction and coercion, (e) false advertisement for

"modelling, acting or dancing" opportunities, (f) peer recruitment, (g) Internet enticement through chat rooms and profile sharing, and (h) use of social media. The emphasis of this report is on categories (g) and (h).

With the rise of new technologies, traffickers are adapting their modus operandi, increasing usage of cyberspace technologies. While adults are more exposed to trafficking through free-standing websites (e.g., escort sites), children and teenagers are more often abused through social media, where traffickers take advantage of lacking awareness of dangers of exploitation – 31% of victims on social media are children or teenagers [37]. At the phase of victim recruitment, the following patterns have been reported and described in more detail below. (a) Posting of deceptive job offers; (b) Direct communication with victims; (c) Threatening.

Deceptive job offers. Advertisements describe the possibility of living a luxurious life or promising jobs in industries such as modelling or entertainment. Until the mid-2000s, such advertisements were most often found on classified webpages or free-standing webpages than on social media but are increasingly replaced by social media, which has grown its share from zero to 51 per cent according to an analysis of court cases [37].

Temporary contracting agencies bring in workers through legitimate means under the auspices of luring people with the promise of work to lead a better life. However, the victims are charged exorbitant fees that they can never pay because they are often never paid for their work. This fee was issued to subjugate and exploit the workers, forcing them to tolerate and endure intolerable situations [37].

Direct communication with victims using the anonymity of online spaces**.** Traffickers make use of the difficulties in identifying the authors writing on social media. Furthermore, **the fast pace of communication** makes social media the ideal tool for traffickers to stay in contact with both victims and customers. In one case reported in [37], the trafficker used multiple online profiles to recruit the victims. The trafficker contacted each victim through **two fake identities**: one was used to write abusive text messages, while the other was used to express understanding and compassion. This technique was instrumental in building trust with the victims.

Also, according to Microsoft. "How Vulnerabilities Increase Child Sex Trafficking Risk." , Caitlin Allen. "The Role of the Internet on Sex Trafficking." [4,30], an essential first step on the side of traffickers is building trust and psychological leverage to exploit their victims. A reported example mentions a girl who was a runaway working for Lawson, and she was initially contacted in the fall of 2008 on myspace.com, where she was promised to be made a 'star' [37]. Family rejection, lack of support systems, and financial challenges each offer heightened opportunities for traffickers to step in and exploit **LGBTQ youth** in particular. Traffickers can target these vulnerabilities to fill the role of a trusted adult [4].

A pattern used by traffickers is to move from open communication channels (e.g., open groups in social media) to encrypted and anonymised services like WhatsApp, aware of surveillance risks [37].

Threatening. The techniques used by Traffickers also include convincing victims to share revealing pictures of themselves under the guise of assessment for modelling job offers, then using the very same photographs to blackmail their victims into further abusive actions. The same can happen after victims have been physically abused and videos of the abuse were recorded [37]. Aside from generating profit in itself, pornography is used as a form of psychological manipulation to keep victims locked in the industry by making them believe they are permanently 'tarnished' since they are now on the Internet forever [30].

Table 3 shows an overview of technologies used in the identified steps related to child trafficking. The recruitment of victims is split **into multiple steps, as explained** in the above section (Approaching and recruitment of victims using digital technologies), and thus is represented in multiple columns.

Table 3. Human Trafficking - Mapping technology with modus operandi stages.

	Identification	Recruitment: Job offers	Recruitment: Direct communication	Recruitment: Threatening	Exploitation	Marketing and Dissemination
Social Media (public)	++	++	++	–	–	+
Messaging Apps and private chats	++	++	++	++	++	++
Video Chat Apps/streams	~	~	–	–	++	++
Official Dating Sites/ chat rooms related to dating	+	~	~	–	–	–
Classified Webpages for advertisement	–	~ * * more up until the mid-2000s	–	–	–	++
Free-standing webpages (created by traffickers)	–	+ * * temporary contracting agencies	–	+ * * temporary contracting agencies	+ * * chat rooms monitored by traffickers	+ * * online bulletin boards, chat rooms, and members-only forums

Legend			
++	+	~	–
Very Frequent	**Frequent**	**Uncommon**	**Never or almost never**

The exploitation of Human Trafficking victims using digital technologies

In general, once recruited, the trafficked children are controlled using a combination of physical, psychological, and sexual abuse and the use of drugs and alcohol to break down their psychological defences. Children are often traded between traffickers and moved between various locations to evade prosecution and further disorientation [31]. Victims are also isolated without access to technology [37].

On the other hand, digital platforms have transformed the patterns of exploitation, like webcams and live streams have created new forms of exploitation and reduced the need for transportation and transfer of victims [38]. Traffickers have coerced victims into establishing rapport with customers in chat rooms monitored by the traffickers [38]. Digital technology can be used as a surveillance tool to trap victims in the industry, e.g., a trafficker was reported to place recording devices on phones to track her every movement [30].

Marketing of human trafficking via digital technologies

On the consumer side, Escort and Massage Services helping as fronts for prostitution and other sexual activities advertise the availability of the trafficked victims using online directories and social media such as Facebook and Twitter. The facilities and their licit and illicit services may be the subject of reviews, discussions, and third-party directories on the Internet. In particular, the clients ("johns") share notes about facilities, escort services, and particular experiences and individuals using online bulletin boards ("john boards") as well as social media, including both private and public forums [31]. Several other sites, including backpage.com, cityvibe.com, eros.com, humaniplex.com, myredbook.com, and sugardaddyforme.com, continue to host ads where underage and trafficked persons may be involved [31]. Open sites are believed to host "john board" bulletin boards and chat rooms where clients of sexual services communicate, including eroticmp.com and theeroticreview.com. Besides the "open" material, there are also password-protected "members only" forums on these and other sites [31].

Detection

Little information on the detection of (child) human trafficking is available. As mentioned, Traffickers are using the anonymity of Online spaces, often communicating directly with victims and switching between multiple channels. A typical pattern used by traffickers is to move from open communication channels (e.g., open groups in social media) to encrypted and anonymised services like WhatsApp, aware of surveillance risks [37], making detection complicated. Further, there is a lack of structure at authorities, needed to identify cases [34]. In addition, most of the populations relevant to the study of human traffickings, such as victims/survivors of trafficking for sexual exploitation, traffickers, or illegal migrants, are part of a "hidden population" (i.e., it is almost impossible to establish a sampling frame and draw a representative sample of the population) [30].

However, two technical solutions were identified that work in the area of detecting human trafficking; both use data mining and AI: Wang, Hao, Andrew Philpot, EH Hovy, and M Latonero. "Data Mining and Integration to Combat Child Trafficking." [31] present a prototype of a law enforcement support system called TraffickBot to automatically compile and correlate information from open sources about trafficking and sexual abuse of women and especially children. The system employs information retrieval, information integration, and NLP technologies to build a data warehouse allowing various visualizations of information for the benefit of law enforcement. Further, W. Chung, E. Mustaine, and D. Zeng. "Criminal Intelligence Surveillance and Monitoring on Social Media: Cases of Cyber-Trafficking." [32] focus on Criminal intelligence surveillance and monitoring on social media for Cases of cyber-trafficking. They present the Cyber-Trafficking Surveillance System (CyTraSS) and provide preliminary findings of using the system to monitor human trafficking social media discussions. However, in reality, detection methods appear to be ineffective. This ineffectiveness is probably due to the limited amount of data available and the wide variety of methods used by criminals (and their high adaptability). Addressing this problem requires coordinated actions to raise public awareness and improve detection methods using the latest advances in AI to detect suspicious online patterns.

4.4. Misinformation

Our society has experienced incredible transformations in recent years, particularly in the area of communication. Today we have vastly more sources of information than past generations. Moreover, we can interact with that information or create our own. This has important implications for discerning between truthful and false information. Technology has not created hoaxes, stereotypes, and false information, but it has made it possible to scale their use (and spreading) enormously and make it easier for anyone to contribute to this issue (sometimes unintentionally or without malicious intent).

The authors Thomas, Kurt, Devdatta Akhawe, Michael Bailey, Dan Boneh, Elie Bursztein, Sunny Consolvo, Nicola Dell, et al. "SoK: Hate, Harassment, and the Changing Landscape of Online Abuse." [5] argue that **researchers on the mission to study disinformation and misinformation** face a dilemma between restricting content and research access. Platforms face pressure to remove fake information promptly, thus making it difficult to quantify such information retroactively. In [5], mainly hate and harassment are analysed, but the authors argue that the **boundary to violent extremism** and disinformation/misinformation is not always clear. An example given is the so-called "Pizzagate", which started as a conspiracy theory that presidential candidate Hillary Clinton was involved in a paedophilia ring that was run out of a Pizza restaurant, which led to violent harassment of the restaurant's staff.

However, despite such effects, the **main intention** behind misinformation campaigns is still to change the perception of as many people as possible (for political, financial, or recreational reasons). **Social media** is playing a fundamental role in the spreading of misinformation. During the 2016 US presidential election campaign, the 20 most frequently discussed fake election stories

generated 8,711,000 shares, reactions, and comments on Facebook [17]. These effects are particularly relevant to **minors**, as younger people, in general, tend more intensely to inform themselves from social media instead of traditional media like TV and newspapers. A 2015 national survey states that, among **younger millennials**, there is a strong tendency to get news from social media [87].

According to Michael B. Robb. "News and America's kids: How young people perceive and are impacted by the news." [88], the most preferred sites for news for teens are Facebook (47%), YouTube (14%), and Twitter (13%). **Social media does appear to play a substantial role in introducing youth to the news**. However, Journell, Wayne. "Unpacking Fake News: An Educator's Guide to Navigating the Media with Students." [87] highlights that these trends do not suggest that youth rely solely on social media for news; both studies found that youth were aware of the need to follow up and go to other sources, particularly for "hard news" about serious social or political topics. Figure 2 shows the leading social media platforms used by minors (between 4 and 15) in 2020 [24].

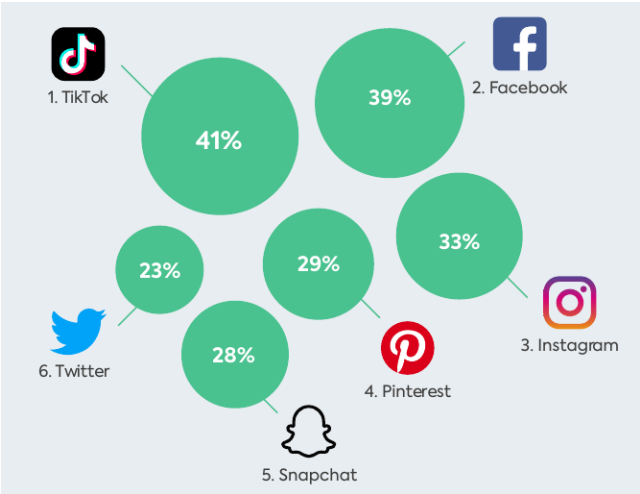

Figure 2. Ranking of social networks according to time of use by children between 4 and 15. Source: [24]

Social media has the advantage of breaking the physical distance barrier among individuals and providing platforms to share, forward, vote, and review. Users are encouraged to participate and discuss online news, which can lead to repercussions and substantial potential political and economic benefits, encouraging malicious entities to create, publish, and spread fake news. Studies in social psychology have also found that the **human ability to detect deception** is around 54% - only slightly better than pure chance [17]. According to a recent study [88], only 44% of **teens** agreed that they could **tell the difference between real and fake news**. In addition, 31% of kids who shared a news story online in the last six months say they shared a story they later found out was wrong or inaccurate.

The very functioning of social networks and their **recommendation algorithms** contribute enormously to this problem [106] [57]. They tend to create information bubbles, where users mainly see content that reaffirms their preconceived ideas, making them very vulnerable to

misinformation. The only way to "burst" this bubble is for the user to consciously decide to seek out new types of content and perspectives. In the broadcasting of disinformation (malicious information), bots or fake accounts also come into play, which disseminates certain types of content in order to trick recommendation algorithms and achieve the virilization of false information.

X. Zhou et al. [18] identify four different categories by which fake news can be detected automatically: (i) The knowledge itself; (ii) the writing style; (iii) the way the information propagates; (iv) and the credibility of the information source. Regarding writing style, recent advances in NLP AI such as GPT-3 (capable of generating incredibly realistic text) make the task of detecting fake news increasingly complicated. Although for the moment, the most catastrophic predictions about the implications of these algorithms have not been fulfilled.

An article by Facebook AI [15] describes efforts taken by Facebook to protect users from harmful content, scaling the work of human experts, and proactively acting before inappropriate content can harm users. Policies implemented by Facebook include adding or warnings and context to content, reducing the distribution of information rated as problematic, or event removing such misinformation. A challenge remains the compelling interplay between human fact-checkers and AI tools, where the former provides high-quality ground truth while the latter brings scalability. Another critical challenge is the multi-modality of misinformation – the same content can be phrased differently, or the format might be switched between text, images, and videos. To address these challenges, Facebook has developed many tools, including SimSearchNet++, ObjectDNA, and LASER.

There are numerous proposals for **detecting fake news** in the academic sphere, but these are mainly based on the writing style. As we have explained above, the latest advances in AI will make the task of detection based on text alone increasingly challenging. In Pérez-Rosas, Verónica, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. “Automatic Detection of Fake News.” [16], automatic detection of fake news is addressed, focusing on online news using ML. In the experiments, accuracies of up to 76% are reported. Also, in Reis, Julio C. S., Andre Correia, Fabricio Murai, Adriano Veloso, Fabricio Benevenuto, and Erik Cambria. “Supervised Learning for Fake News Detection.” [17], the challenge of fake news detection is addressed; the authors provide several new features that help train AI models for such tasks, and they discuss how detection approaches can be employed in practice. However, new forms of misinformation have appeared in image, audio, and video format, the so-called **DeepFakes**. Currently, they do not pose a significant threat because the technology is not yet mature enough. Nevertheless, in a few years, these hoaxes will undoubtedly be almost undetectable. There are proposals to detect DeepFakes automatically [26], but it remains a very challenging problem.

In recent years, social networks have shifted their moderation methods almost exclusively to AI and reduced human moderators, contrary to expert recommendations. AI algorithms are promising, but they still have a long way to go before being reliable, safe, and ethical. This is mainly due to the very subjective and nuanced nature of the problem. These AI algorithms are still a very effective tool to

make the work of human moderators much more straightforward. Unfortunately, when Covid-19 forced these workers out of their offices, large tech companies began to rely more on AI-based moderation. As a result, more fake news and hate speech got past the filters [27].

Another approach to combat misinformation on the Internet is to improve **users' critical thinking** and help them to detect misinformation or fake news better. One approach is to **teach children and teenagers to spot misinformation** from an early age: For example, Chang, Yoo Kyung, Ioana Literat, Charlotte Price, Joseph I Eisman, Jonathan Gardner, Amy Chapman, and Azsaneé Truss. "News Literacy Education in a Polarised Political Climate: How Games Can Teach Youth to Spot Misinformation." [89] designed, implemented, and evaluated a game about fake news to test its potential to enhance news literacy skills in educational settings. The game is a card game—set in a fictional narrative environment—designed to teach middle and high school students strategies to identify misinformation. The authors of the study and game found that students could transfer in-game learning to real-world contexts.

Addressing the issue of misinformation requires coordinated actions from many perspectives, including:

- Improve users' critical thinking and help them with (perhaps AI-based) tools to contrast the information they receive.
- Make technology platforms and social networks use effective methods to control misinformation and hold them accountable for their role in its dissemination.
- Create international frameworks where governments, scientists, and journalists collaborate to refute false information that is disseminated.

5. Survey to RAYUELA's Experts

During the development of this deliverable, a questionnaire was sent to members of the RAYUELA consortium who could provide helpful information on the cybercrimes analysed and the role that technology plays in them. Specifically, the questionnaire was sent to LEAs:

- PLV – Policía Local de Valencia (Spain)
- PJ – Polícia Judiciária (Portugal)
- PSNI – Police Service of Northern Ireland (Northern Ireland)
- EPBG – Estonian Police and Border Guard Board (Estonia)

And partners with key positions in the education sector:

- EA – Ellinogermaniki Agogi (Greece)
- UCLL – University College Leuven-Limburg (Belgium)

We believe that this first-hand information from relevant actors enriches the contributions already made in the report. However, we must be aware that such information may be biased because these members operate in specific locations and serve a significantly smaller population than can be analysed in large-scale studies.

5.1. Cyberbullying

- **Question:** *Regarding the crime of Cyberbullying: Which age ranges are more likely to suffer it?*

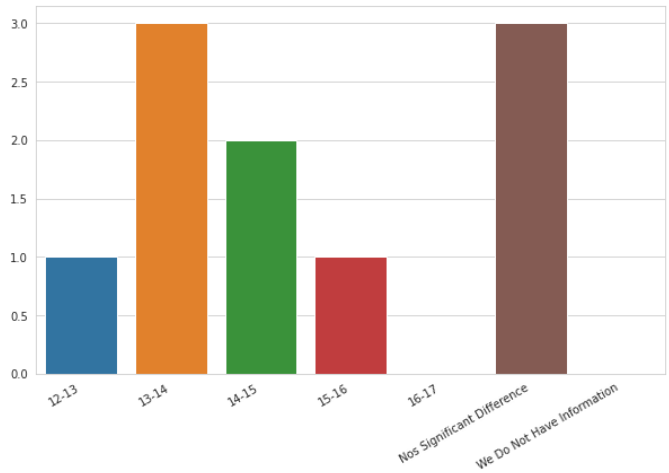

Figure 3. RAYUELA's experts answers - Age ranges most likely of victims of cyberbullying.

Figure 3 shows a summary of the answers provided by the consortium members. We can observe that most of the partners agree that there is a peak in the age range of 13 to 14 years old. This coincides with the data provided in 2017 by the ANAR Foundation, which works with minors facing these problems in Spain [33].

- **Question:** *Regarding the crime of Cyberbullying: Which gender is more likely to suffer it?*

Figure 4 shows a summary of the answers provided by the consortium members. Most of them indicate that there is not a significant difference. According to data from the ANAR Foundation [33], there is no significant difference in *bullying* depending on the gender of the victim. However, in the case of *cyberbullying*, women suffer it in a much higher proportion (Women: 65.6%; Men: 34.4%).

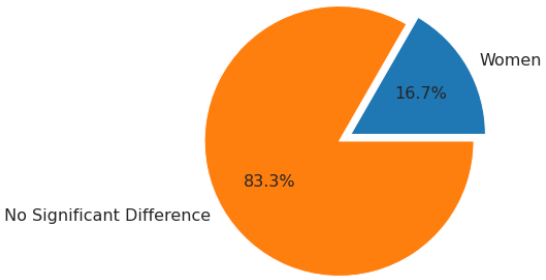

Figure 4. RAYUELA's experts answers - Gender more likely to be victims of cyberbullying.

- **Question:** *Regarding the crime of Cyberbullying: Which apps, webs, or social media are most commonly used to commit this crime?*

Figure 5 shows a summary of the answers provided by the consortium members. Most of them indicate that WhatsApp is the most common media to commit cyberbullying. According to data from the ANAR Foundation [33], WhatsApp is the most used method for cyberbullying in Spain. It was used in 76% of the reported cases, and other social networks were used in 48%. Moreover, in 92.6% of the cases, a smartphone was used to commit the harassment.

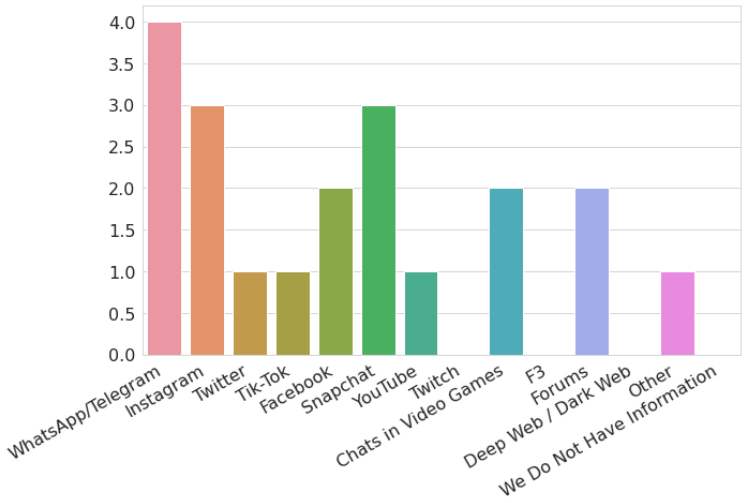

Figure 5. RAYUELA's experts answers - Most used applications/social media to commit cyberbullying.

Additional Comments:

- Some of the partners pointed out that the smartphone is by far the most used tool for cyberbullying.
- Many cases have also seen that the victim and the offender had a (more or less) close relationship. However, usually, the offenders try to act anonymously. The anonymity increases the power of the offender over the victim.
- Direct insults or offensive words are the most common type of cyberbullying, followed by threats.

5.2. Online Grooming

- **Question:** *Regarding the crime of Online Grooming: Which age ranges are more likely to suffer it?*

Figure 6 shows a summary of the answers provided by the consortium members. We observed in their responses that there does not seem to be a specific age group more prone to suffer online grooming. According to data from the ANAR Foundation [82], the highest number of reported cases in Spain was in the 13 to 15 age group (<=12 years old: 20,83%; 13-15 years old: 45,83%; >=16 years old: 33,33%).

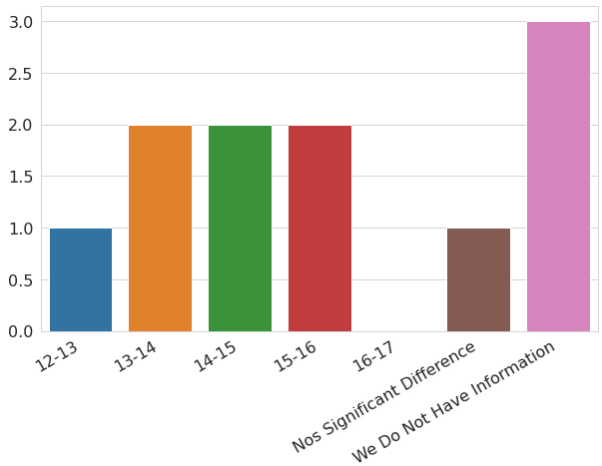

Figure 6. RAYUELA's experts answers - Age ranges most likely of victims of online grooming.

- **Question:** *Regarding the crime of Online Grooming: Which age gender is more likely to suffer it?*

Figure 7 shows a summary of the answers provided by the consortium members. Among the partners who have information on the subject, half of them indicate that women are more likely to suffer from online grooming, and the other half indicates that there is no significant difference. According to data from the ANAR Foundation [82], in 69,57% of the reported cases in Spain, the victim was a woman.

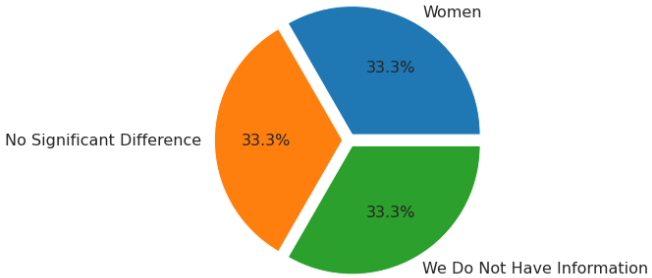

Figure 7. RAYUELA's experts answers - Gender more likely to be victims of online grooming.

- **Question:** *Regarding the crime of Online Grooming: Which apps, webs, or social media are most commonly used to commit this crime?*

Figure 8 shows a summary of the answers provided by the consortium members. Most of the partners who have information on the subject indicate that Instagram is the most used social network for committing this crime, which is consistent with UK data provided by the NSPCC [80].

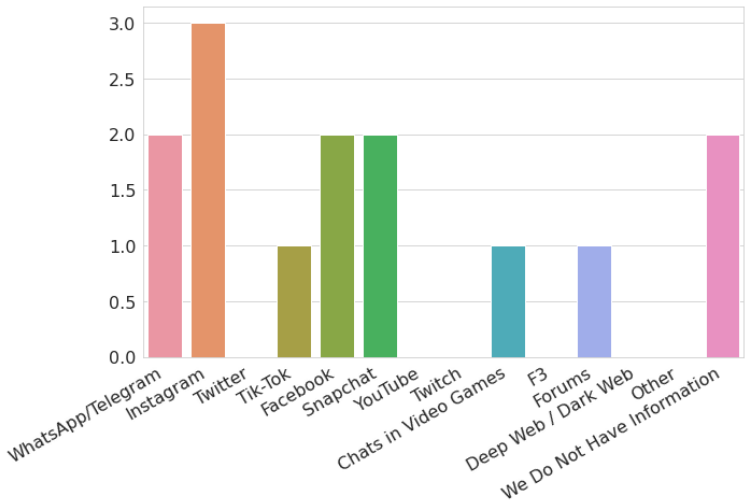

Figure 8. RAYUELA's experts answers - Most used applications/social media to commit online grooming.

Additional Comments:

- In most online grooming cases, the offender identifies a victim in online forums, social networks, or video game chats. Offenders adapt to the media most commonly used by minors. Once the offender has gained the child's trust, they begin to use sexualised language and request sexual photographs/videos. Once the criminal has obtained some data or compromising material from the minor, he/she uses extortion to obtain more material. Sometimes the offender offers gifts to the minor to make them feel indebted.
- Victims often come from families with fewer opportunities. The children are frequently alone and look for friends on the Internet.

- On numerous occasions, it has been reported that the criminal poses as a person close to the victim's age.
- In recent years, an increase in the use of video game platforms/chats to contact minors has been detected.
- In some countries such as Colombia and the Philippines, there are links between online grooming and human trafficking. There, traffickers pose as troubled teenagers in social media just like the teenagers they target. However, this trend has not yet been observed significantly in Europe.

5.3. Human Trafficking

- **Question:** *Regarding the crime of Human Trafficking: Which age ranges are more likely to suffer it?*

Figure 9 shows a summary of the answers provided by the consortium members. Most of the consortium members do not have direct information related to human trafficking. According to data from the ANAR Foundation [82], in the cases of prostitution of minors reported in Spain, the most common age range is between 13 and 17, with a lower number of cases involving children under 12. However, we must keep in mind that not all cases of human trafficking lead to prostitution and that not all cases of prostitution occur through a human trafficking network.

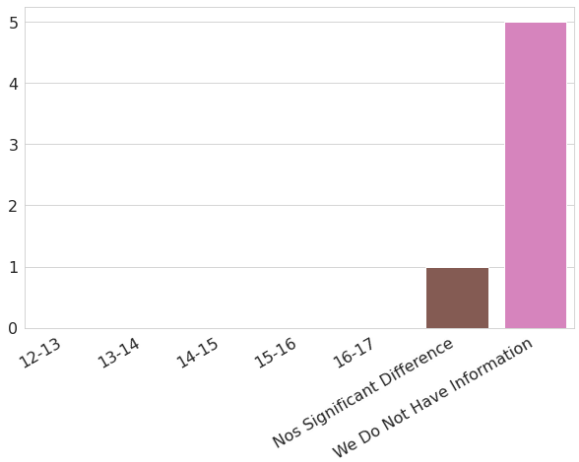

Figure 9. RAYUELA's experts answers - Age ranges most likely of victims of human trafficking.

- **Question:** *Regarding the crime of Human Trafficking: Which age gender is more likely to suffer it?*

Figure 10 shows a summary of the answers provided by the consortium members. Those consortium members who have information on this issue indicate that women are more likely to be victims of human trafficking. According to data from the ANAR Foundation [82], in 80% of the cases of prostitution reported in Spain, the victim was a woman. A study on human trafficking in the United States [30] indicates that 96% of sex trafficking cases involve women and girls. However, once again,

we must keep in mind that not all cases of human trafficking lead to prostitution and that not all cases of prostitution occur through a human trafficking network.

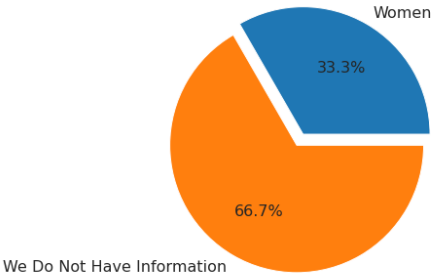

Figure 10. RAYUELA's experts answers - Gender more likely to be victims of human trafficking.

- **Question:** *Regarding the crime of Human Trafficking: Which apps, webs, or social media are most commonly used to commit this crime?*

Figure 11 shows a summary of the answers provided by the consortium members. Most of the consortium members do not have direct information related to human trafficking. Responses from those who do have information indicate that victims are contacted through social networks and forums.

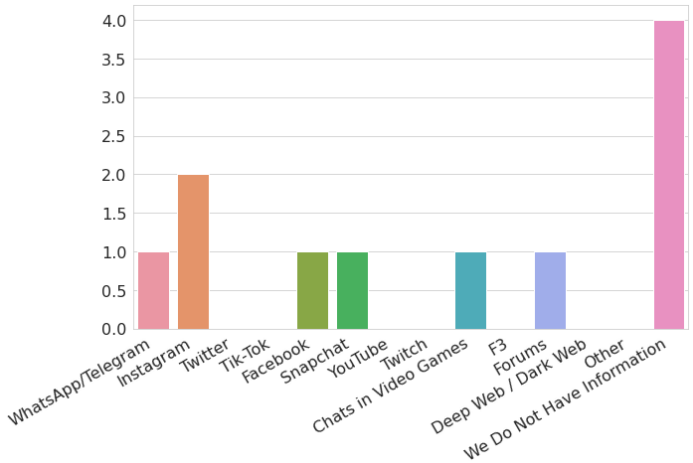

Figure 11. RAYUELA's experts answers - Most used applications/social media to commit human trafficking.

Additional Comments:

- In countries like Greece, where large numbers of war refugees pass through, human trafficking mafias are very prevalent.

5.4. Misinformation

- **Question:** *Regarding Misinformation: Which age ranges are more likely to suffer it?*

Figure 12 shows a summary of the answers provided by the consortium members. Most consortium members agree that there is no significant difference in age for being a victim of misinformation.

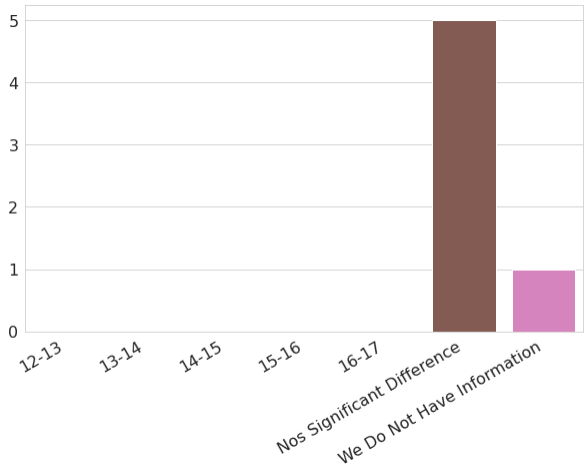

Figure 12. RAYUELA's experts answers - Age ranges most likely be victims of misinformation.

- **Question:** *Regarding Misinformation: Which age gender is more likely to suffer it?*

Figure 13 shows a summary of the answers provided by the consortium members. Those members who have some information on this agree that there is no significant difference in gender for being a victim of misinformation.

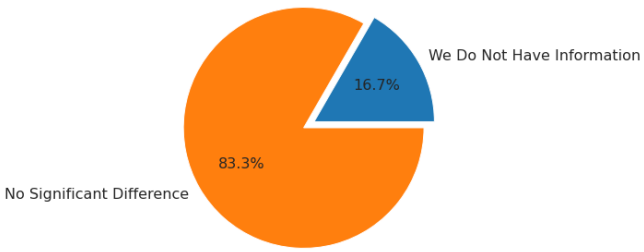

Figure 13. RAYUELA's experts answers - Gender more likely to be victims of misinformation.

- **Question:** *Regarding Misinformation: Which apps, webs, or social media are most commonly used to commit it?*

Figure 14 shows a summary of the answers provided by the consortium members. The responses indicate that the most prominent social networks have the highest number of misinformation cases (e.g., Facebook, TikTok, Twitter, Instagram, WhatsApp). Forums also seem to be relevant in this issue.

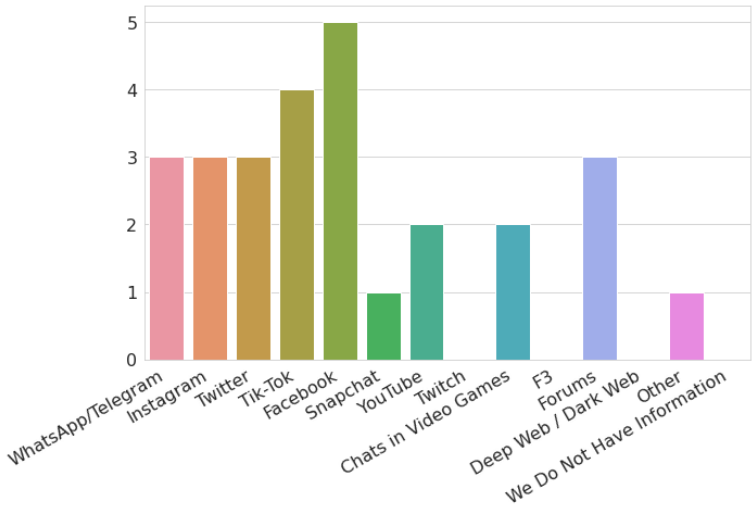

Figure 14. RAYUELA's experts answers - Most used applications/social media to commit misinformation.

Additional Comments:

- Facebook seems to be the social network that spreads more misinformation, according to some reports. However, this social network is not one of the most used by European minors.
- The age of the people who suffer from misinformation depends mainly on the topic of the information.
- Under the pandemic situation due to Covid-19, a significant increase in fake news across all platforms.
- Some studies indicate that misinformation is transmitted in greater quantity through smartphones (compared to computers) because these devices invite us to "think less and click more".
- The use of fake profiles and bots to spread false information is very common.

6. Survey to Minors

During the development of this report, we have been conducting surveys to children in European schools. This survey has two simultaneous objectives: (i) to obtain first-hand information on the use of technology by European children and the human/demographic factors that may affect online victimisation in the cybercrimes considered; (ii) to test and extend the network of schools collaborating with RAYUELA and to motivate them to prototype the RAYUELA serious game in the future.

We believe that this first-hand information enriches the contributions already made in the report. However, here we will only analise the results obtained in Spain, as this is the only country where the survey process has been completed to date. Future deliverables and scientific publications will also present the data collected in more European countries. In addition, in this report we will focus on analysing the survey questions related to technological topics. Since social and psychological aspects will be analysed further by WP1 in the open deliverable 1.7.

All the questions contained in the survey can be found in Appendix A. The survey was answered by 840 children studying in (public and private) Spanish schools, aged between 12 and 17 (Mean=14.56, SD=0.9. 48.8% of participants identified themselves as male, 44% as female, and 3% as non-binary.

6.1. Most Used Applications

- **Question:** *Order from MOST to LEAST the applications you use the most:*
  - *WhatsApp / Telegram*
  - *Instagram*
  - *TikTok*
  - *Snapchat*
  - *Twitter*
  - *Facebook*
  - *Online Video Games (with audio/text chats)*
  - *Dating Applications (Tinder, Grinder, etc.)*
  - *YouTube*
  - *Twitch*

Through this question we attempted to find out where European teenagers spend most of their time when they are on the Internet. To do so, participants had to rank the listed applications in order of highest to lowest usage. In Figure 15, we can observe the frequency with which each application was placed in the first position of the ranking, where Instragram seems to be the clear winner, ahead of TikTok.

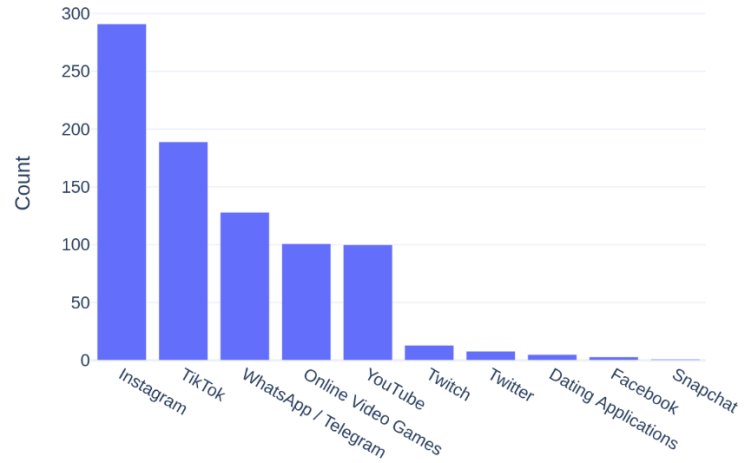

Figure 15. RAYUELA survey to minors: Most used applications

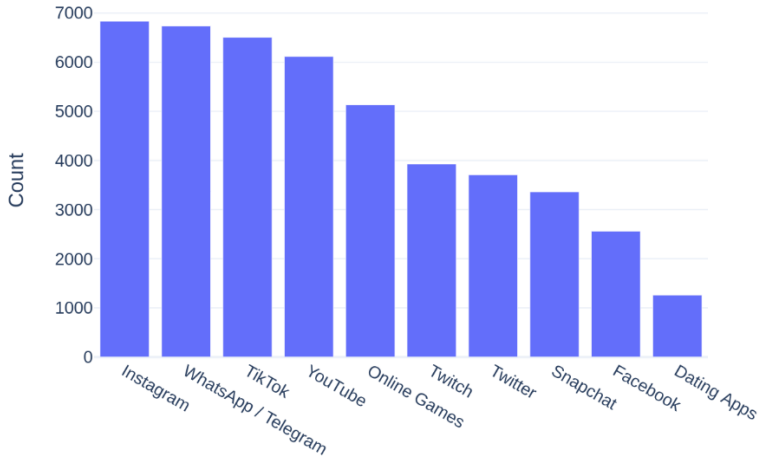

Figure 16. RAYUELA survey to minors: Most used applications (Accumulated)

To get a complete picture of which applications are most used, we can accumulate all rakings positions by giving a numerical value to each position (i.e., position 1 is given a value of 10, position 2 is given a value of 9, ..., position 10 is given a value of 1). In Figure 16, we observe the result of performing this cumulative operation, where Instagram is still in the lead, closely followed by WhatsApp / Telegram and TikTok. Dating applications are the least used applications on the list by a significant margin.

Figure 17 shows the distributions of each application according to the number of times they appear in each position of the rakings. In these figures, the X-axis stands for the positions in the rakings from 1 to 10, and the Y-axis represents the number of times the applications appear in that position. The figure of "Online Games" is the most homogeneous of all. This suggests that almost all minors play online games, although in a widely varying degree regarding the amount of time spent on it.

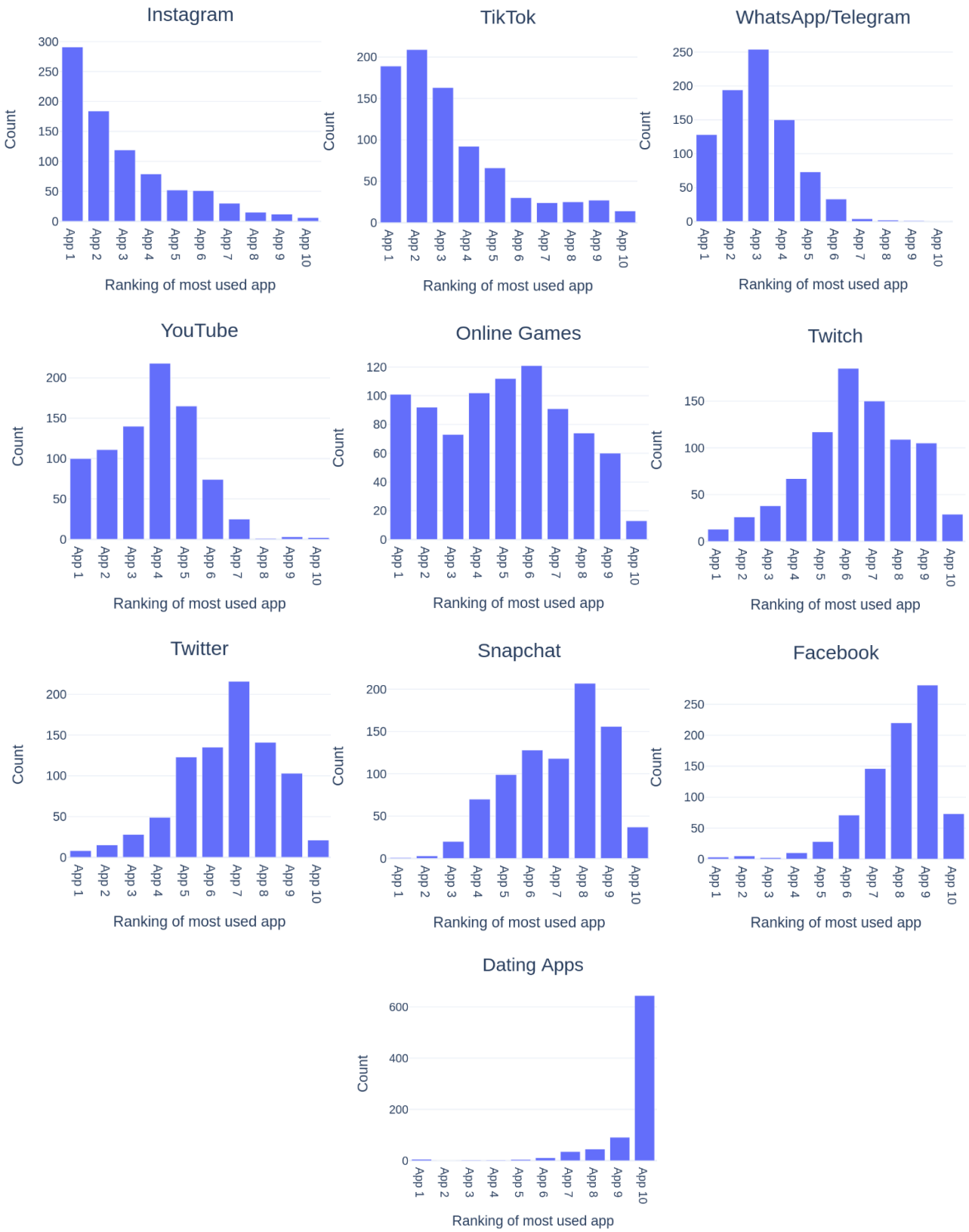

Figure 17. RAYUELA survey to minors: Distributions of the considered Apps' positions in the ranking of most used apps. 'App 1' stands for the first position in the ranking, and 'App 10' for the last position.

6.2. Hours spent on the Internet for leisure

- **Question:** *Enter the number of hours you spend per day on the Internet for entertainment (social networking, gaming, etc.) on average on a **weekday**.*
  - ☐ *Less than 1 hour per day*
  - ☐ *Between 1 and 2 hours per day*
  - ☐ *Between 3 and 4 hours per day*
  - ☐ *More than 4 hours per day*
  - ☐ *Don't know / No answer*
- **Question:** *Enter the number of hours you spend per day on the Internet for entertainment (social networking, gaming, etc.) on average on a **weekend** day.*
  - ☐ *Less than 1 hour per day*
  - ☐ *Between 1 and 2 hours per day*
  - ☐ *Between 3 and 4 hours per day*
  - ☐ *More than 4 hours per day*
  - ☐ *Don't know / No answer*

Through these questions we attempted to better understand the amount of time European teenagers spend on the Internet for entertainment purposes. To get a more granular answer, we asked for the number of hours on weekdays and weekends, guided by the intuition that in these two scenarios the answers are very different. The difference between Figure 18 and Figure 19 confirms this conjecture, showing that the differences in hours spent on the Internet are significant. At the weekend, the number of teenagers who spend more than 4 hours on the Internet for entertainment grows considerably.

Figure 18 shows that most adolescents spend between 1 and 4 hours (on the Internet for entertainment) per day on weekdays. However, a significant number also spend more than 4 hours per day. Very few teenagers spend less than 1 hour a day. This distribution shifts considerably towards more hours when it comes to the weekend. In Figure 19, we see how at the weekend most teenagers spend more than 4 hours a day on the Internet. Figure 20 shows an accumulation of both options (weekdays and weekends) where it can be seen that spending more than 4 hours on the Internet per day is the mode of the distribution, i.e., the option most chosen by teenagers.

According to WP1's research, spending a large number of hours on the Internet, or the addition to it, is a relevant risk factor in some of the cybercrimes considered.

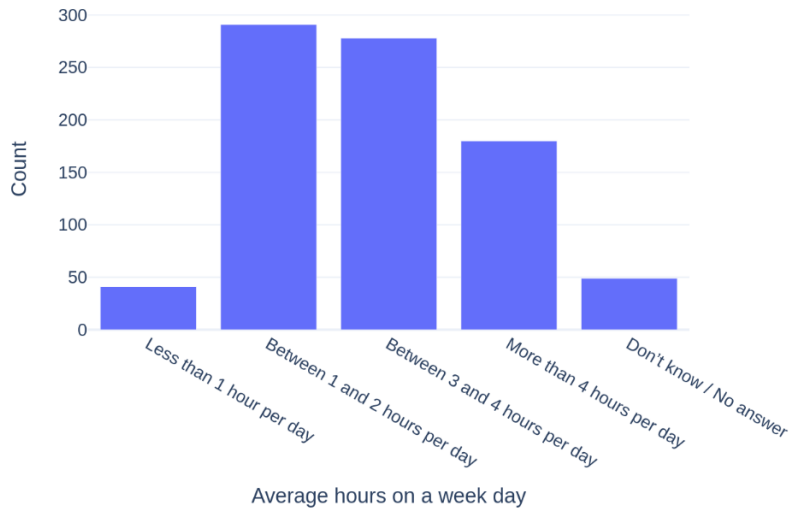

Figure 18. RAYUELA survey to minors: Average hours spend on the Internet for leisure, on a weekday

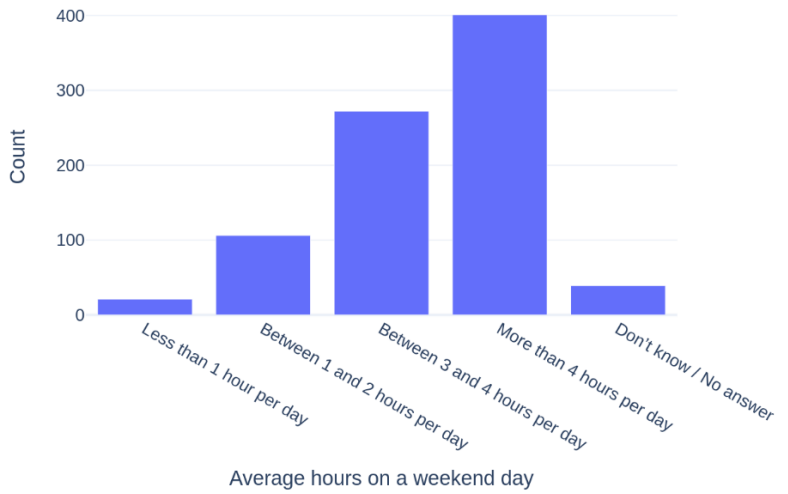

Figure 19. RAYUELA survey to minors: Average hours spend on the Internet for leisure, on a weekend day

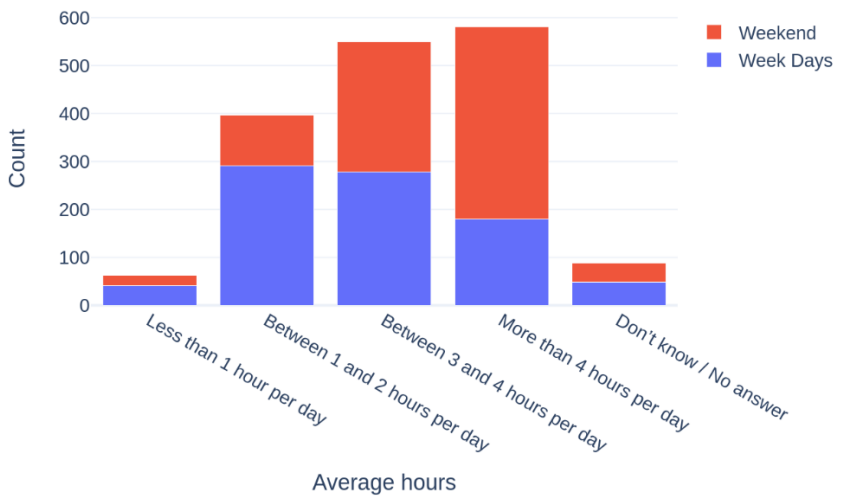

Figure 20. RAYUELA survey to minors: Average hours spend on the Internet for leisure, both on weekdays and weekend days

6.3. Connected Devices

- **Question:** *Please indicate which of the following devices you use the most. 1 being "I don't use it", and 5 being "I use it a lot".*
  - *Smartphone*
  - *Tablet*
  - *PC*
  - *Game Console (PlayStation, Xbox, Nintendo, etc.)*
  - *Smartwatch*
  - *Smart Speaker (Amazon Echo Alexa, Google Home, Apple HomePod, etc.)*
  - *Smart Toys (i.e., toys that connect to smartphoes or the Internet)*
- **Question:** *Please indicate which of the following devices are your favourites. 1 being "I don't like it", and 5 being "I love it". If you do not use any of the devices you can indicate this by marking the option 0 "I don't use it".*
  - *Smartphone*
  - *Tablet*
  - *PC*
  - *Game Console (PlayStation, Xbox, Nintendo, etc.)*
  - *Smartwatch*
  - *Smart Speaker (Amazon Echo Alexa, Google Home, Apple HomePod, etc.)*
  - *Smart Toys (i.e., toys that connect to smartphoes or the Internet)*
- **Question:** *From the point of view of cybersecurity and personal data privacy, which devices do you think are the most secure to use? Please rate the devices, with 1 being "Not secure", and 5 being "Very secure".*
  - *Smartphone*
  - *Tablet*
  - *PC*
  - *Game Console (PlayStation, Xbox, Nintendo, etc.)*
  - *Smartwatch*
  - *Smart Speaker (Amazon Echo Alexa, Google Home, Apple HomePod, etc.)*
  - *Smart Toys (i.e., toys that connect to smartphoes or the Internet)*

Through these questions we have attempted to identify which connected devices are most used by teenagers, as well as their perceptions of the security and privacy of these devices. To do so, participants were asked to indicate with a score from 1 to 5 how much they use the listed devices, how much they like using these devices, and to rate them from their perceptions of the security and privacy the devices offer.

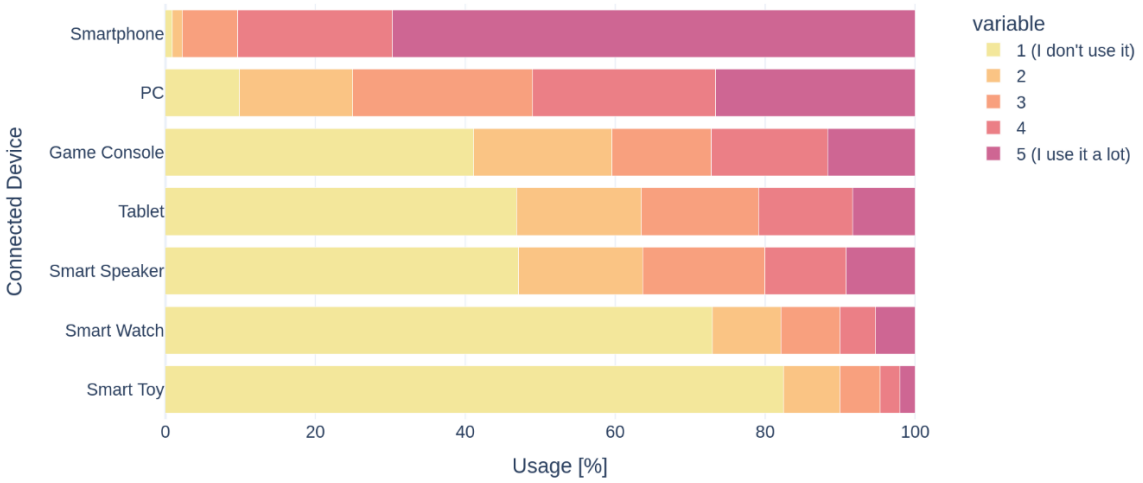

Figure 21. RAYUELA survey to minors: Most used connected devices

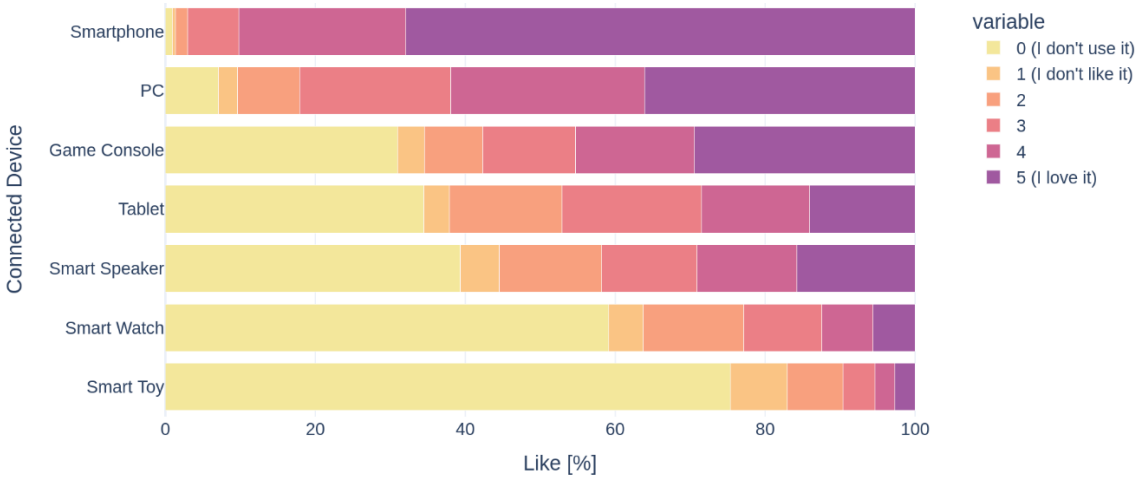

Figure 22. RAYUELA survey to minors: Most favourite connected devices

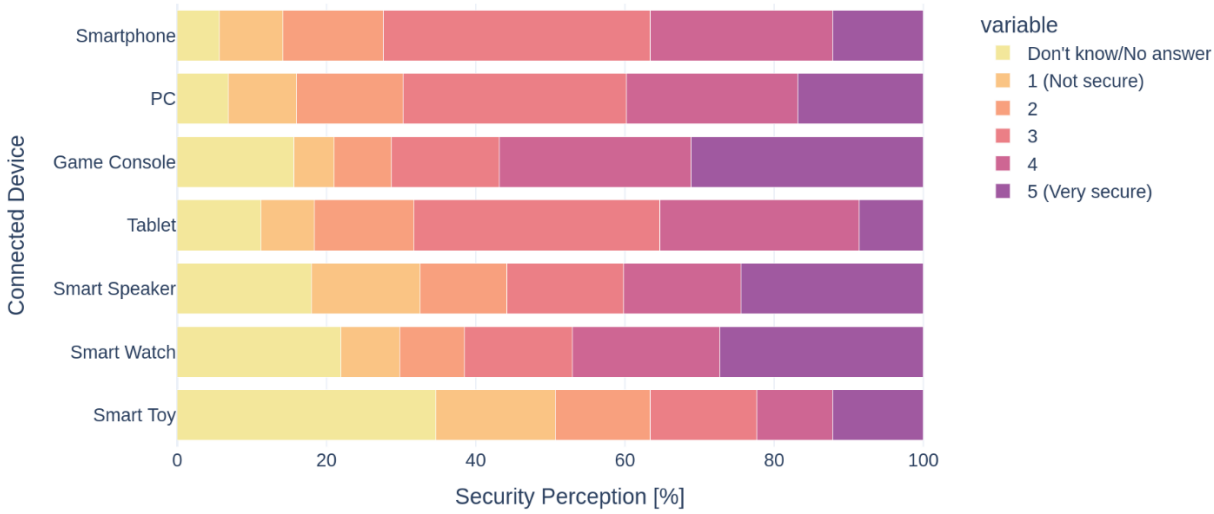

Figure 23. RAYUELA survey to minors: Security and privacy perception of connected devices

Figure 21 shows the results of the question of how much they use each device. We note that the smartphone is by far the most used device, followed by the PC. Subsequently, with similar ratings we find the game console, the tablet, and the Smart Speaker. The Smart Watch and Smart Toys are the least used devices. Figure 22 shows participants' favourite devices. We see that the results are similar to Figure 21, with the exception of the game console which is placed closer to the PC. Indicating that many teenagers would like to use the game console more than they do.

Figure 23 shows participants' perceptions of the devices in terms of security and privacy. We observe that the distributions are now much more heterogeneous. The game console seems to be the best perceived device from a cybersecurity standpoint, but with little difference to the rest, except for Smart Toys, which do seem to have a poor perception of their security and privacy, although this device is also where most participants indicated "Don't know / No answer".

To obtain a more holistic view of these perceptions in the following figures we have plotted the cumulative of all scores, giving each score its equivalent numerical value (i.e., rating a 5 gives a value of 5, rating a 4 gives a value of 4, etc.). Figure 24 shows the aggregate for each device comparing the variables of use (X-axis) and perception of security and privacy (Y-axis). The size of each circle in the figure indicates the amount of people responding “Don’t know/No answer”, as a measure of uncertainty. Equivalently, Figure 25 compares the aggregates of the variables favourite devices and perceptions of security and privacy.

In both figures, we note that despite small differences, the perception of security of all devices is similar, with the exception of Smart Toys, which have the lowest score in usage, favorability and perception of security and privacy. These figures seem to suggest that young people's perception of security in connected devices is quite homogeneous, they do not perceive that there are devices much more insecure than others.

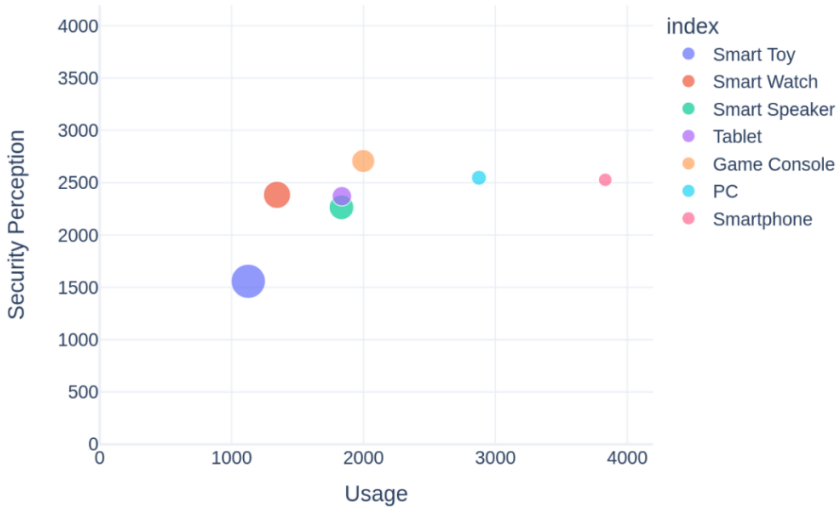

Figure 24. RAYUELA survey to minors: Confronting the amount of usage and the perception of security, on each connected device. The size of the circles indicates the amount of people responding “Don’t know/No answer”, as a measure of uncertainty.

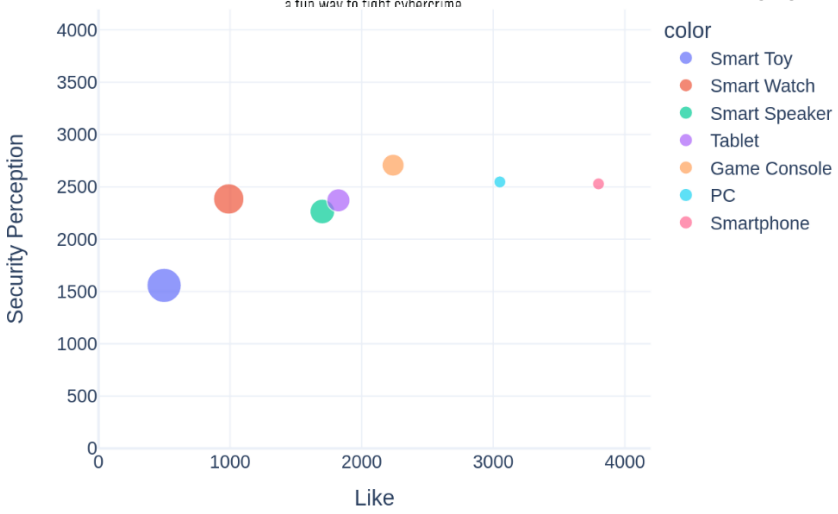

Figure 25. RAYUELA survey to minors: Confronting the amount of liking and the perception of security, on each connected device. The size of the circles indicates the amount of people responding "0 (I don't use it)" as a measure of uncertainty.

6.4. Tor Browser

- **Question:** *Do you know the Tor browser?*
  - ☐ *Yes*
  - ☐ *No*
  - ☐ *Don't know / No answer*
- **Question:** *If you answered 'Yes' to the above question, have you ever used it? What have you ever used it for?*
  - ☐ *[Open Answer]*

Through these questions we tried to determine what percentage of the participants know or have used the Tor browser. This also helps us to infer which participants are more knowledgeable about the world of cybersecurity, in which the Tor browser is an important tool. In addition, the Tor browser can also be used to access the *Darknet*, where illegal content can be found and shared.

Figure 26 shows that about 13% of the participants have heard of the Tor browser. Figure 27 shows how this percentage drops significantly when asked if they have ever used Tor. Among the participants who claim to have used Tor, the most common responses were out of curiosity, to search for information about drugs, or to view content not available in their country or not accessible from the usual web.

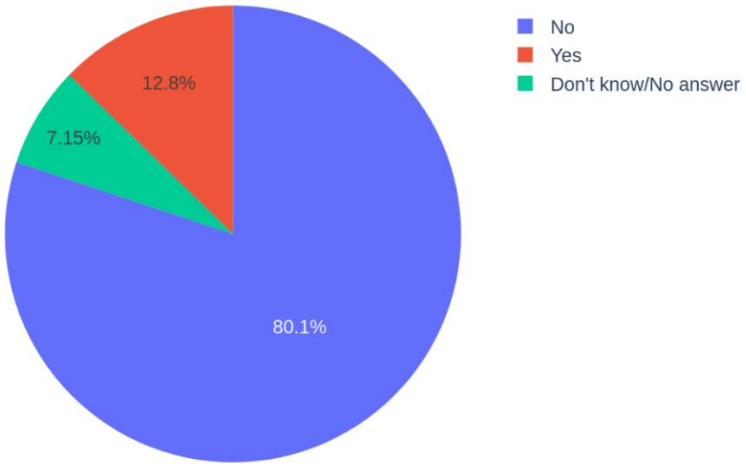

Figure 26. RAYUELA survey to minors: Percentage of participants knowing the Tor browser.

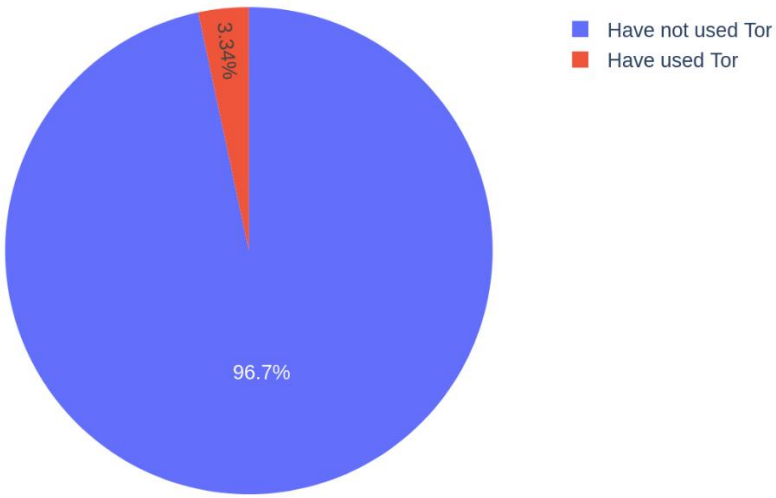

Figure 27. RAYUELA survey to minors: Percentage of participants who have used the Tor browser.

7. Discussion

7.1. Role of Technology and AI to Improve the Online Experience of Minors

Nowadays, we are encountering new generations that not only were “born digital”, but also born embedded in social networks and in a digital society that we are still structuring. On many occasions, through these digital media, minors can find themselves in situations for which they (and even their caretakers) are not prepared. It is not yet clear whether this has more benefits or detriments. Technology (particularly AI) is a powerful tool to address complex problems in new and sophisticated ways. Our generation must direct AI towards social goals such as helping new generations empower themselves and making them feel protected in the (digital) world. In concrete, Technology can improve children's online experience, identifying three main fields of action: **prevention**, **detection**, and **mitigation**.

Prevention

The most immediate preventive measure is to raise awareness among children and parents of the potential online issues they may encounter. There are concrete examples in the literature that address the cybercrimes considered: cyberbullying [92], online grooming [93] [94], human trafficking [95], and misinformation [96] [97]. There are also initiatives promoted by the European Union [90] and UNICEF [91], which have the same awareness-raising objective. In this area of action, it is vital for governments and social networks (and other online platforms) to work closely together. Here is where awareness campaigns or more innovative solutions such as video games (like RAYUELA) can directly and significantly impact. Consequently, awareness campaigns must also go hand in hand with learning about good practices in privacy and cybersecurity, for instance:

- Avoid using the same nicknames on all platforms.
- Avoid sharing personal data or material, especially when having a public account.
- Avoid accepting everybody to the user's friends list.
- Improve users' critical thinking skills and help them with tools (perhaps AI-based) that make their online experience more pleasant and less prone to risk.

Focusing on the use of technology and AI for prevention, digital platforms should impose further restrictions on account creation to avoid bots or fake accounts (e.g., require an official document ID). Traditional techniques for detecting fake accounts are generally not very effective [46], as they only analyse account characteristics to look for anomalies (e.g., profile information, number of followers, ...). AI techniques can help detect fake profiles [49] and suspicious behaviours. Another potential use of AI is detecting risky profiles or behaviours to focus awareness and detection campaigns on those users. With the latest advances in chatbots and language processing technology, virtual assistants could help children, and beginners correctly configure their security privacy settings and, ideally, introduce improved mechanisms to avoid uploading sensitive content.

Detection

The analysis conducted throughout Section 4 highlights several techniques and mechanisms based on technology or AI to detect the most prevalent crimes. However, the main limitation of these detection techniques is that they address the problem from only one standpoint. Specifically, automatic detection is based exclusively on text analysis or targeting child pornography by searching only pre-existing databases. In practice, the cybercrimes described in the preceding sections are tremendously complex, and criminals are constantly adapting their methods so that no single-perspective solution can handle them in a clear-cut way.

However, the latest advances in AI give us a glimmer of hope about the possible applications in improving children's online experience. For example, novel language processing methods invite us to think that we will be able to understand with great precision the context and nuances based on the semantic analysis of the conversation. Besides, the field of computer vision (and self-supervised learning) has also made spectacular advances in recent years, so we will soon be able to improve our systems for detecting abuse material. Finally, advances in Deep Learning and the analysis of (online) human behaviour pave the ground for automatically detecting risk profiles that protect the most vulnerable in the digital world. Naturally, the practical success of these solutions requires their integration to address all possible fronts (e.g., risk profile detection [100] + metadata [101] [102] [103] + Computer Vision [102] + NLP [104] [105] + etc.). Unfortunately, the field of AI is not yet mature enough to be a complete fail-safe end-to-end solution (e.g., adversarial attacks on neural networks, model bias, lack of transparency).

Mitigation

Finally, prompt and effective countermeasures are required once a specific cybercrime has already occurred. Thus, mitigation actions require the same agility with which illicit or harmful content is generated on the Internet to trigger a proportional response when denouncing and reporting those felonies. In particular, a closer and more agile collaboration between digital platforms and LEAs is mandatory. This closer relationship should result in shorter intervention times, and children should find it easier to communicate when experiencing problems of this type.

AI should also play a role at this stage of mitigation, for example, the proposal promoted by UNICEF called *SomeBuddy* [98]: When a minor reports an incident, such as cyberbullying, the AI-based system automatically analyses the case using NLP and prepares a "first aid kit" for the victim. Lawyers and psychologists always supervise it, but this tool allows them to attend to more minors and (potentially) more effectively. Another prominent example that UNICEF powered is the *Milli chatbot* [99]: This conversational system based on NLP is designed to help teens express themselves and learn about mental health. Other studies, such as Chelmis, Charalampos, and Mengfan Yao. “Minority Report: Cyberbullying Prediction on Instagram.” [10], focus on developing an automated

approach to reducing the reaction time between detection and intervention, and it has been validated on Instagram data.

In summary, recent advances in AI make us firmly believe that AI will play an essential role in developing tools to prevent, detect, and mitigate online problems children face. Nevertheless, some critical problems still need to be solved to be a handy tool in the real world (e.g., the privacy of children's data, transparency of algorithmic decisions, security against adversarial attacks, mitigating model and data biases, and ethics). While these issues are being addressed, current and near-future techniques can serve as essential tools for professionals to assist minors faster and effectively.

7.2. Takeaways for RAYUELA's Serious Game

The analysis carried out in the previous sections focused on the design and implementation of the video adventures in the RAYUELA game. In particular, it will help us to define scenarios and scripts consistent with how cybercrime happens in the real world. In the following, we will outline the main highlights that we can extract for the game in each cybercrime.

Cyberbullying

- This crime is committed mainly in **messaging applications** and **social networks** and is adapted to those most used at the given time. The most widely used electronic device for cyberbullying is the **smartphone**, and the most widely used platform is **WhatsApp**.
- A major **risk factor** that facilitates cyberbullying is the **underestimation of sharing personal information** on social networks and using technology without correct knowledge of it and its dangers. Even so, in most cases, the offender had a (more or less) **close relationship** with the victim (e.g., they attend the same class/school).
- According to data from ANAR Foundation [33] and RAYUELA expert partners, risk peaks at **ages** between 13 and 14 years old. However, age is not a limiting factor; cyberbullying still occurs in other age groups, although somewhat to a lesser extent.
- According to data from the ANAR Foundation [33] and some of the RAYUELA expert partners, it seems that **girls** suffer cyberbullying more frequently. However, in the case of bullying (no cyberbullying), the percentages between genders are similar.
- Some studies show a significant correlation between cyberbullying victimization and online grooming [83].

Online Grooming

- The victim **approach** phase usually occurs mainly in messaging applications, social networks, and gaming chats. The sexual **harassment** phase usually occurs in messaging applications or

private social network chats. The **dissemination** of the material obtained by the criminal usually occurs through P2P networks and darknet forums.

- A major **factor** that facilitates online grooming is underestimating **the risk behind sharing personal information** on social networks and using technology without correct knowledge of it and its dangers. Especially if the minor has a **public account** on social networks or he/she accepts everybody to their friends' list. Besides, they use the **same nickname** on all digital platforms (thus making it easier to locate them).
- It is common for victims to be **vulnerable children**, minors in care or with disabilities, or those at risk of economic and social exclusion.
- According to data from the ANAR Foundation [82] and RAYUELA expert partners, risk peaks at **ages** between 13 to 15 years old. However, age is not a limiting factor; it still occurs in other age groups, although somewhat to a lesser extent.
- According to data from the ANAR Foundation [82] and some of the RAYUELA expert partners, it seems that **girls** suffer online grooming more frequently.

Human Trafficking

- Human trafficking is a very **heterogeneous** crime in terms of the means it uses. As in online grooming, a typical pattern is **approaching** the victim through social media and messaging apps and then moving the conversation to private chats. Websites for advertisements or those created by the traffickers are also frequently used.
- The most common **risk factors** are lack of a stable social and economic environment, underestimation of sharing personal information on social media [4, 30], and the search for job offers (usually abroad) that enable the victim to get out of his or her situation of exclusion quickly (e.g., model, dancer, nanny, actress) [37].
- The electronic devices most commonly used for the online exploitation of the victims are smartphones and webcams.
- **LGBTQ+** youth, due to frequent family rejection and lack of social support systems, have a higher risk of becoming victims [4].
- According to data from the ANAR Foundation [82], in the cases of prostitution of minors reported in Spain, the most common **age** range is between 13 and 17, with a lower number of cases involving children under 12. However, the number of reported cases is relatively small.
- According to various studies [82] [30], most human trafficking and prostitution victims are **women**.

Misinformation

- Misinformation is most of the time transmitted through social networks and messaging apps. The recommendation algorithms (tend to create bubbles of information), bots, and fake accounts contribute to this issue significantly.
- Often, real-life elements mix with false ones to create confusing and powerful pieces of misinformation. Also, satire or parody information is complicated to differentiate from misinformation.
- The most significant **risk factor** is having a low critical attitude towards the information received online and solely receiving this information from social networks and forums.

8. Conclusions

This deliverable summarises the most relevant research findings on the impact of technology and the technological threats associated with the cybercrimes considered in the project.

The central insight we can draw from this deliverable is establishing which technologies (e.g., social networks, messaging applications) are used (and how) in the cybercrimes under consideration. The research methodology applied to achieve this overall goal combines conducting a review of the available literature on the addressed topics (Section 4) with conducting surveys to experts, such as LEAs and education institutions within the RAYUELA consortium (Section 5), and minors (Section 6).

In the case of cyberbullying, the analysis of the scientific literature shows that Twitter has attracted much interest from the research community on this topic. However, the experts, based on their experience, points out WhatsApp, Instagram, and Snapchat as the most used applications to commit cyberbullying. In the case of misinformation, it is interesting that data presented in Figure 2 and the opinion of the experts agree upon highlighting TikTok and Facebook as the most important platforms. In the case of online grooming, based on the analysis of the literature, criminals target and contact the victim through social networks and gaming chats and then move the conversation to private chats and encrypted messaging applications. In this case, the experts points out Instagram as the most used application to commit online grooming based on their own experience. Human trafficking is more heterogeneous than the others, although similar to online grooming. The results of the survey to the expert in this case shows that they are not very familiar with this cybercrime. Both in online grooming and in human trafficking, the dissemination of sensitive material obtained by criminals is usually done through P2P networks and darknet forums.

Regarding online habits and use of technology by minors, based on the survey, most of the consulted adolescents spend between 1 and 4 hours online per day on weekdays, a significant number also spend more than 4 hours per day, and this distribution shifts considerably towards more hours when it comes to the weekend. The analyzed reports shows that the COVID-19 pandemic has increased remarkably the time minors spend online. Regarding devices, both the reports and the survey points something that seems to be crystal-clear: the smartphone is the preferred device by minor to be online. Regarding applications, based on the survey, Instagram, WhatsApp/Telegram, and TikTok are the most used ones, which we think that meet the reality for this age range. Based on the data from the analysis of the literature, the experts, and the survey regarding applications, it could be concluded that the applications most used to commit cybercrimes are the most popular ones between the target age ranges, what somehow confirms the hypothesis that criminals know very well the online habits of their victims and go looking for them wherever they are.

This invites all the stakeholders involved in the online ecosystem to put even more efforts to protect minors' online in those platforms that are or start becoming popular between them. In addition,

this analysis will have also significant relevance when building the adventures of the RAYUELA game, since it will allow developing them coherently with real-world cases.

Finally, based on the literature reviewed and the information gathered, we are convinced that AI will play a key role in developing technologies to improve minors' online experience. However, some significant issues must be addressed to be a helpful tool in real-world cases, being ethical aspects of capital importance in this sense, as discussed extensively in Section 7.

9. Bibliography

[1] Hasse, Alexa, Sandra Cortesi, Andres Lombana Bermudez, and Urs Gasser. "Youth and Cyberbullying: Another Look," October 2019. https://dash.harvard.edu/handle/1/41672537.

[2] Microsoft. "Meet the New Anti-Grooming Tool from Microsoft, Thorn, and Our Partners." Thorn (blog), February 11, 2020. https://www.thorn.org/blog/what-is-project-artemis-thorn-microsoft-grooming/.

[3] Microsoft. "Online Grooming: What It Is, How It Happens, and How to Defend Children." Thorn (blog), June 15, 2020. https://www.thorn.org/blog/online-grooming-what-it-is-how-it-happens-and-how-to-defend-children/.

[4] Microsoft. "How Vulnerabilities Increase Child Sex Trafficking Risk." Thorn (blog), January 17, 2020. https://www.thorn.org/blog/how-vulnerabilities-increase-child-sex-trafficking-risk/.

[5] Thomas, Kurt, Devdatta Akhawe, Michael Bailey, Dan Boneh, Elie Bursztein, Sunny Consolvo, Nicola Dell, et al. "SoK: Hate, Harassment, and the Changing Landscape of Online Abuse." In 2021 IEEE Symposium on Security and Privacy (SP), 247–67. San Francisco, CA, USA: IEEE, 2021. https://doi.org/10.1109/SP40001.2021.00028.

[6] Silva, Yasin N., Deborah L. Hall, and Christopher Rich. "BullyBlocker: Toward an Interdisciplinary Approach to Identify Cyberbullying." Social Network Analysis and Mining 8, no. 1 (December 2018): 18. https://doi.org/10.1007/s13278-018-0496-z.

[7] Chatzakou, Despoina, Nicolas Kourtellis, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Athena Vakali. "Mean Birds: Detecting Aggression and Bullying on Twitter." In Proceedings of the 2017 ACM on Web Science Conference, 13–22. Troy New York USA: ACM, 2017. https://doi.org/10.1145/3091478.3091487.

[8] Yao, Mengfan, Charalampos Chelmis, and Daphney-Stavroula Zois. "Cyberbullying Ends Here: Towards Robust Detection of Cyberbullying in Social Media." In The World Wide Web Conference on - WWW '19, 3427–33. San Francisco, CA, USA: ACM Press, 2019. https://doi.org/10.1145/3308558.3313462.

[9] Cheng, Lu, Jundong Li, Yasin N. Silva, Deborah L. Hall, and Huan Liu. "XBully: Cyberbullying Detection within a Multi-Modal Context." In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, 339–47. Melbourne VIC Australia: ACM, 2019. https://doi.org/10.1145/3289600.3291037.

[10] Chelmis, Charalampos, and Mengfan Yao. "Minority Report: Cyberbullying Prediction on Instagram." In Proceedings of the 10th ACM Conference on Web Science - WebSci '19, 37–45. Boston, Massachusetts, USA: ACM Press, 2019. https://doi.org/10.1145/3292522.3326024.

[11] Rosa, H., N. Pereira, R. Ribeiro, P.C. Ferreira, J.P. Carvalho, S. Oliveira, L. Coheur, P. Paulino, A.M. Veiga Simão, and I. Trancoso. "Automatic Cyberbullying Detection: A Systematic Review." Computers in Human Behaviour 93 (April 2019): 333–45. https://doi.org/10.1016/j.chb.2018.12.021.

[12] Balakrishnan, Vimala, Shahzaib Khan, and Hamid R. Arabnia. "Improving cyberbullying detection using Twitter users' psychological features and machine learning." Computers & Security 90 (2020): 101710.

[13] Hasse, Alexa, Sandra Clio Cortesi, Andres Lombana, and Urs Gasser. "Youth and Artificial Intelligence: Where We Stand." SSRN Electronic Journal, 2019. https://doi.org/10.2139/ssrn.3385718.

[14] Smith, Peter K, Jess Mahdavi, Manuel Carvalho, Sonja Fisher, Shanette Russell, and Neil Tippett. "Cyberbullying: Its Nature and Impact in Secondary School Pupils." Journal of Child Psychology and Psychiatry 49, no. 4 (2008): 376–85.

[15] Facebook. "Here's how we're using AI to help detect misinformation." Facebook AI (blog), November 19, 2020. https://ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misinformation/.

[16] Pérez-Rosas, Verónica, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. "Automatic Detection of Fake News." In Proceedings of the 27th International Conference on Computational Linguistics, 3391–3401. Santa Fe, New Mexico, USA: Association for Computational Linguistics, 2018. https://aclanthology.org/C18-1287.

[17] Reis, Julio C. S., Andre Correia, Fabricio Murai, Adriano Veloso, Fabricio Benevenuto, and Erik Cambria. "Supervised Learning for Fake News Detection." IEEE Intelligent Systems 34, no. 2 (March 2019): 76–81. https://doi.org/10.1109/MIS.2019.2899143.

[18] Zhou, Xinyi, and Reza Zafarani. "A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities." ACM Computing Surveys 53, no. 5 (October 15, 2020): 1–40. https://doi.org/10.1145/3395046.

[19] Facebook. "Child protection on the Internet." Safety Center. Accessed September 9, 2021. https://www.facebook.com/safety/onlinechildprotection.

[20] Madigan, Sheri, Anh Ly, Christina L. Rash, Joris Van Ouytsel, and Jeff R. Temple. "Prevalence of Multiple Forms of Sexting Behaviour Among Youth: A Systematic Review and Meta-Analysis." JAMA Pediatrics 172, no. 4 (April 1, 2018): 327–35. https://doi.org/10.1001/jamapediatrics.2017.5314.

[21] Susman-Peña, Tara, Mehri Druckman, and Nina Oduro. "Fighting Misinformation Digital Media Literacy". The Teaching Company. USA, 2020.

[22] Wardle, Claire, and Hossein Derakhshan. "Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making." Council of Europe 27 (2017).

[23] Smahel, David, Hana Machackova, Giovanna Mascheroni, Lenka Dedkova, Elisabeth Staksrud, Kjartan Ólafsson, Sonia Livingstone, and Uwe Hasebrink. "EU Kids Online 2020: Survey Results from 19 Countries." London School of Economics and Political Science, 2020. http://doi.org/10.21953/lse.47fdeqj01ofo.

[24] Qustodio. "Qustodio annual report on children's digital habits." Qustodio (blog), April 7, 2021. https://www.qustodio.com/en/2021/04/07/qustodio-annual-report-on-childrens-digital-habits/.

[25] Chelmis, Charalampos, and Daphney-Stavroula Zois. "Dynamic, Incremental, and Continuous Detection of Cyberbullying in Online Social Media." ACM Transactions on the Web 15, no. 3 (May 13, 2021): 1–33. https://doi.org/10.1145/3448014.

[26] Tolosana, Ruben, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, and Javier Ortega-Garcia. "Deepfakes and beyond: A Survey of Face Manipulation and Fake Detection." Information Fusion 64 (December 1, 2020): 131–48. https://doi.org/10.1016/j.inffus.2020.06.014.

[27] Scott, Mark, and Kayali, Laura. "What Happened When Humans Stopped Managing Social Media Content," POLITICO. October 21, 2020. https://www.politico.eu/article/facebook-content-moderation-automation/.

[28] Vanderschaaf, Victoria. "Spotlight on: How the Internet Facilitates Underage Victimization in Human Trafficking." Child. Legal Rts. J. 34 (2013): 135.

[29] Goodman, Miriam, and Julie Laurence. "Child Trafficking Victims and the State Courts." In A Guide to Human Trafficking for State Courts, 77–88. Human Trafficking and the State Courts Collabourative, Denver (CO), 2014.

[30] Caitlin Allen. "The Role of the Internet on Sex Trafficking." International Observatory of Human Rights. March 7, 2019. https://observatoryihr.org/blog/the-role-of-the-internet-on-sex-trafficking/.

[31] Wang, Hao, Andrew Philpot, EH Hovy, and M Latonero. "Data Mining and Integration to Combat Child Trafficking." Retrieved from Carnegie Mellon University, School of Computer Science, 2014. http://www.cs.cmu.edu/~hovy/papers/12dgo-trafficking.pdf.

[32] W. Chung, E. Mustaine, and D. Zeng. "Criminal Intelligence Surveillance and Monitoring on Social Media: Cases of Cyber-Trafficking." In 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), 191–93, 2017. https://doi.org/10.1109/ISI.2017.8004908.

[33] ANAR Foundation. "III Estudio Sobre Acoso Escolar y Ciberbullying Según Los Afectados." ANAR Foundation, 2017. https://www.anar.org/wp-content/uploads/2018/09/III-Estudio-sobre-acoso-escolar-y-ciberbullying-seg%C3%BAn-los-afectados.pdf.

[34] European Fundamental Rights Agency. "Child Trafficking in the European Union: Challenges, Perspectives and Good Practices." LU: Publications Office of the European Union, 2009. https://data.europa.eu/doi/10.2811/12279.

[35] Group of Experts on Action against Trafficking of Human Beings (GRETA). "GRETA's 6th General Report," Council of the European Union. March 2017. https://ec.europa.eu/anti-trafficking/sites/default/files/greta_2017_7_web_6gr_en.pdf.pdf.

[36] Latonero, Mark. "Human Trafficking Online: The Role of Social Networking Sites and Online Classifieds." University of Southern California, School for Communication & Journalism. September 1, 2011. http://dx.doi.org/10.2139/ssrn.2045851

[37] United Nations Office on Drugs and Crime. "Traffickers Use of the Internet; Digital Hunting Fields." United Nations Office on Drugs and Crime, 119–28. Global Report on Trafficking in Persons. United Nations, 2021. https://doi.org/10.18356/9789210051958c009.

[38] UNICEF and Innocenti Research Centre. "Child Trafficking in Europe: A Broad Vision to Put Children First." The United Nations Children's Fund (UNICEF), 2008.

[39] Calvo-Morata, Antonio, Manuel Freire-Moran, Ivan Martinez-Ortiz, and Baltasar Fernandez-Manjon. "Applicability of a Cyberbullying Videogame as a Teacher Tool: Comparing Teachers and Educational Sciences Students." IEEE Access 7 (2019): 55841–50. https://doi.org/10.1109/ACCESS.2019.2913573.

[40] Salawu, Semiu, Yulan He, and Joanna Lumsden. "Approaches to Automated Detection of Cyberbullying: A Survey." IEEE Transactions on Affective Computing 11, no. 1 (January 1, 2020): 3–24. https://doi.org/10.1109/TAFFC.2017.2761757.

[41] Al-Garadi, Mohammed Ali, Mohammad Rashid Hussain, Nawsher Khan, Ghulam Murtaza, Henry Friday Nweke, Ihsan Ali, Ghulam Mujtaba, Haruna Chiroma, Hasan Ali Khattak, and Abdullah Gani. "Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges." IEEE Access 7 (2019): 70701–18. https://doi.org/10.1109/ACCESS.2019.2918354.

[42] A. S. Srinath, H. Johnson, G. G. Dagher, and M. Long. "BullyNet: Unmasking Cyberbullies on Social Networks." IEEE Transactions on Computational Social Systems 8, no. 2 (April 2021): 332–44. https://doi.org/10.1109/TCSS.2021.3049232.

[43] W. M. Al-Rahmi, N. Yahaya, M. M. Alamri, N. A. Aljarboa, Y. B. Kamin, and F. A. Moafa. "A Model of Factors Affecting Cyber Bullying Behaviours Among University Students." IEEE Access 7 (2019): 2978–85. https://doi.org/10.1109/ACCESS.2018.2881292.

[44] Chatzakou, Despoina, Ilias Leontiadis, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, Athena Vakali, and Nicolas Kourtellis. "Detecting Cyberbullying and Cyberaggression in Social Media." ACM Transactions on the Web 13, no. 3 (November 18, 2019): 1–51. https://doi.org/10.1145/3343484.

[45] Mladenović, Miljana, Vera Ošmjanski, and Staša Vujičić Stanković. "Cyber-Aggression, Cyberbullying, and Cyber-Grooming: A Survey and Research Challenges." ACM Computing Surveys 54, no. 1 (April 2021): 1–42. https://doi.org/10.1145/3424246.

[46] N. Singh, T. Sharma, A. Thakral, and T. Choudhury. "Detection of Fake Profile in Online Social Networks Using Machine Learning." In 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), 231–34, 2018. https://doi.org/10.1109/ICACCE.2018.8441713.

[47] Tahmasbi, Nargess, and Elham Rastegari. "A Socio-Contextual Approach in Automated Detection of Public Cyberbullying on Twitter." ACM Transactions on Social Computing 1, no. 4 (December 21, 2018): 1–22. https://doi.org/10.1145/3290838.

[48] Soni, Devin, and Vivek K. Singh. "See No Evil, Hear No Evil: Audio-Visual-Textual Cyberbullying Detection." Proceedings of the ACM on Human-Computer Interaction 2, no. CSCW (November 2018): 1–26. https://doi.org/10.1145/3274433.

[49] S. Khaled, N. El-Tazi, and H. M. O. Mokhtar. "Detecting Fake Accounts on Social Media." In 2018 IEEE International Conference on Big Data (Big Data), 3672–81, 2018. https://doi.org/10.1109/BigData.2018.8621913.

[50] McInroy, Lauren B., and Faye Mishna. "Cyberbullying on Online Gaming Platforms for Children and Youth." Child and Adolescent Social Work Journal 34, no. 6 (December 1, 2017): 597–607. https://doi.org/10.1007/s10560-017-0498-0.

[51] Aizenkot, Dana. "Social Networking and Online Self-Disclosure as Predictors of Cyberbullying Victimization among Children and Youth." Children and Youth Services Review 119 (December 1, 2020): 105695. https://doi.org/10.1016/j.childyouth.2020.105695.

[52] Lowry, Paul Benjamin, Gregory D. Moody, and Sutirtha Chatterjee. "Using IT Design to Prevent Cyberbullying." Journal of Management Information Systems 34, no. 3 (July 3, 2017): 863–901. https://doi.org/10.1080/07421222.2017.1373012.

[53] Zhou, Yingfan, and Rosta Farzan. "Designing to Stop Live Streaming Cyberbullying: A Case Study of Twitch Live Streaming Platform." In C&T '21: Proceedings of the 10th International Conference on Communities & Technologies - Wicked Problems in the Age of Tech, 138–50. Seattle WA USA: ACM, 2021. https://doi.org/10.1145/3461564.3461574.

[54] Apple. "CSAM Detection – Technical Summary." Child Safety, Apple. August, 2021. https://www.apple.com/child-safety/pdf/CSAM_Detection_Technical_Summary.pdf.

[55] ANAR Foundation. "Informe Anual Teléfono/Chat ANAR – En Tiempos de Covid-19 – Año 2020." ANAR Foundation, 2021. https://www.anar.org/wp-content/uploads/2021/07/Informe-ANAR-COVID_Definitivo.pdf

[56] Craig, Wendy, Meyran Boniel-Nissim, Nathan King, Sophie D. Walsh, Maartje Boer, Peter D. Donnelly, Yossi Harel-Fisch, et al. "Social Media Use and Cyber-Bullying: A Cross-National Analysis of Young People in 42 Countries." Journal of Adolescent Health 66, no. 6 (June 2020): S100–108. https://doi.org/10.1016/j.jadohealth.2020.03.006.

[57] Whittaker, Joe, Seán Looney, Alastair Reed, and Fabio Votta. "Recommender Systems and the Amplification of Extremist Content." Internet Policy Review 10, no. 2 (June 30, 2021). https://doi.org/10.14763/2021.2.1565.

[58] Kwan, Irene, Kelly Dickson, Michelle Richardson, Wendy MacDowall, Helen Burchett, Claire Stansfield, Ginny Brunton, Katy Sutcliffe, and James Thomas. "Cyberbullying and Children and Young People's Mental Health: A Systematic Map of Systematic Reviews." Cyberpsychology, Behaviour, and Social Networking 23, no. 2 (February 1, 2020): 72–82. https://doi.org/10.1089/cyber.2019.0370.

[59] Kaluarachchi, Chintha, Matthew Warren, and Frank Jiang. "Review: Responsible Use of Technology to Combat Cyberbullying among Adolescents." Australasian Journal of Information Systems 24 (June 8, 2020). https://doi.org/10.3127/ajis.v24i0.2791.

[60] Milosevic, Tijana, and Marko Vladisavljevic. "Norwegian Children's Perceptions of Effectiveness of Social Media Companies' Cyberbullying Policies: An Exploratory Study." Journal of Children and Media 14, no. 1 (January 2, 2020): 74–90. https://doi.org/10.1080/17482798.2019.1695219.

[61] Kloess, Juliane A., Catherine E. Hamilton-Giachritsis, and Anthony R. Beech. "Offense Processes of Online Sexual Grooming and Abuse of Children Via Internet Communication Platforms." Sexual Abuse 31, no. 1 (February 2019): 73–96. https://doi.org/10.1177/1079063217720927.

[62] Apple. "Child Safety." Apple. Updated on September 3, 2021. https://www.apple.com/child-safety/.

[63] O'Brien, Jennifer E., and Wen Li. "The Role of the Internet in the Grooming, Exploitation, and Exit of United States Domestic Minor Sex Trafficking Victims." Journal of Children and Media 14, no. 2 (April 2, 2020): 187–203. https://doi.org/10.1080/17482798.2019.1688668.

[64] Council of Europe. "Protection of Children against Sexual Exploitation and Sexual Abuse," 2012. ISBN 978-92-871-7572-4. https://rm.coe.int/protection-of-children-against-sexual-exploitation-and-sexual-abuse/1680794e97

[65] Mladenović, Miljana, Vera Ošmjanski, and Staša Vujičić Stanković. "Cyber-Aggression, Cyberbullying, and Cyber-Grooming: A Survey and Research Challenges." ACM Computing Surveys 54, no. 1 (April 2021): 1–42. https://doi.org/10.1145/3424246.

[66] Ngejane, C.H, G Mabuza-Hocquet, J.H.P Eloff, and S Lefophane. "Mitigating Online Sexual Grooming Cybercrime on Social Media Using Machine Learning: A Desktop Survey." In 2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (IcABCD), 1–6. Durban, South Africa: IEEE, 2018. https://doi.org/10.1109/ICABCD.2018.8465413.

[67] Bours, Patrick, and Halvor Kulsrud. "Detection of Cyber Grooming in Online Conversation." In 2019 IEEE International Workshop on Information Forensics and Security (WIFS), 1–6. Delft, Netherlands: IEEE, 2019. https://doi.org/10.1109/WIFS47025.2019.9035090.

[68] Zuo, Zheming, Jie Li, Philip Anderson, Longzhi Yang, and Nitin Naik. "Grooming Detection Using Fuzzy-Rough Feature Selection and Text Classification." In 2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1–8. Rio de Janeiro: IEEE, 2018. https://doi.org/10.1109/FUZZ-IEEE.2018.8491591.

[69] Santisteban, Patricia de, Joana del Hoyo, Miguel Ángel Alcázar-Córcoles, and Manuel Gámez-Guadix. "Progression, Maintenance, and Feedback of Online Child Sexual Grooming: A Qualitative Analysis of Online Predators." Child Abuse & Neglect 80 (June 2018): 203–15. https://doi.org/10.1016/j.chiabu.2018.03.026.

[70] Gámez-Guadix, Manuel, Carmen Almendros, Esther Calvete, and Patricia De Santisteban. "Persuasion Strategies and Sexual Solicitations and Interactions in Online Sexual Grooming of Adolescents: Modelling Direct and Indirect Pathways." Journal of Adolescence 63 (February 2018): 11–18. https://doi.org/10.1016/j.adolescence.2017.12.002.

[71] Google "Fighting Child Sexual Abuse Online." Google, Protecting Children. https://protectingchildren.google/intl/en/.

[72] Winters, Georgia M., Leah E. Kaylor, and Elizabeth L. Jeglic. "Sexual Offenders Contacting Children Online: An Examination of Transcripts of Sexual Grooming." Journal of Sexual Aggression 23, no. 1 (January 2, 2017): 62–76. https://doi.org/10.1080/13552600.2016.1271146.

[73] Los Angeles Criminal Lawyer "Children and Grooming / Online Predators." Child Crime Prevention & Safety Center. 2020. https://childsafety.losangelescriminallawyer.pro/children-and-grooming-online-predators.html.

[74] Hamilton-Giachritsis, Catherine, Elly Hanson, Helen Whittle, Filipa Alves-Costa, and Anthony Beech. "Technology Assisted Child Sexual Abuse in the UK: Young People's Views on the Impact of Online Sexual Abuse." Children and Youth Services Review 119 (December 2020): 105451. https://doi.org/10.1016/j.childyouth.2020.105451.

[75] Townsend, Mark. "Victims in 84% of Online Grooming Cases Are Girls." The Guardian. March 28, 2021. http://www.theguardian.com/society/2021/mar/28/victims-in-84-of-online-grooming-cases-are-girls.

[76] Chan, Tommy K.H., Christy M.K. Cheung, and Zach W.Y. Lee. "Cyberbullying on Social Networking Sites: A Literature Review and Future Research Directions." Information & Management 58, no. 2 (March 1, 2021): 103411. https://doi.org/10.1016/j.im.2020.103411.

[77] University of KwaZulu-Natal, Howard College Campus, South Africa, Simangele Mkhize, and Nirmala Gopal. "Cyberbullying Perpetration: Children and Youth at Risk of Victimization during Covid-19 Lockdown." International Journal of Criminology and Sociology 10 (April 30, 2021): 525–37. https://doi.org/10.6000/1929-4409.2021.10.61.

[78] Khateeb, Haider M al-, and Gregory Epiphaniou. "How Technology Can Mitigate and Counteract Cyber-Stalking and Online Grooming." Computer Fraud & Security 2016, no. 1 (January 2016): 14–18. https://doi.org/10.1016/S1361-3723(16)30008-2.

[79] Lobe, B., Velicu, A., Staksrud, E., Chaudron, S. and Di Gioia, R. "How Children (10-18) Experienced Online Risks during the Covid-19 Lockdown: Spring 2020: Key Findings from Surveying Families in 11 European Countries." European Commission. Joint Research Centre. Publications Office of the European Union, 2021. https://data.europa.eu/doi/10.2760/562534.

[80] National Society for the Prevention of Cruelty to Children (NSPCC). "Instagram Most Recorded Platform Used in Child Grooming Crimes during Lockdown." NSPCC. 2020. http://www.nspcc.org.uk/about-us/news-opinion/2020/instagram-grooming-crimes-children-lockdown/

[81] National Society for the Prevention of Cruelty to Children (NSPCC). "What Parents Need to Know about Sexual Grooming." NSPCC. Updated in 2020. https://www.nspcc.org.uk/what-is-child-abuse/types-of-abuse/grooming/.

[82] ANAR Foundation. "Abuso Sexual en la Infancia y la Adolescencia Según los Afectados y su Evolución en España (2008-2019)." ANAR Foundation, 2020. https://www.anar.org/wp-content/uploads/2021/02/Estudio-ANAR-abuso-sexual-infancia-adolescencia-240221-1.pdf

[83] Gámez-Guadix, Manuel, and Estibaliz Mateos-Pérez. "Longitudinal and Reciprocal Relationships between Sexting, Online Sexual Solicitations, and Cyberbullying among Minors." Computers in Human Behaviour 94 (May 1, 2019): 70–76. https://doi.org/10.1016/j.chb.2019.01.004.

[84] Santisteban, Patricia de, and Manuel Gámez-Guadix. "Estrategias de Persuasión En Grooming Online de Menores: Un Análisis Cualitativo Con Agresores En Prisión." Psychosocial Intervention 26, no. 3 (2017): 139–46. https://doi.org/10.1016/j.psi.2017.02.001.

[85] European Parliament. "Parliament Adopts Temporary Rules to Detect Child Sexual Abuse Online." Press Releases. European Parliament. June 7, 2021. https://www.europarl.europa.eu/news/en/press-room/20210701IPR07503/parliament-adopts-temporary-rules-to-detect-child-sexual-abuse-online.

[86] European Police Office (Europol). "Internet Organised Crime Threat Assessment: (IOCTA) 2016." LU: Publications Office, 2016. https://data.europa.eu/doi/10.2813/275589.

[87] Journell, Wayne. "Unpacking Fake News: An Educator's Guide to Navigating the Media with Students." Teachers College Press, 2019.

[88] Michael B. Robb. "News and America's kids: How young people perceive and are impacted by the news." San Francisco, CA: Common Sense. 2017.

[89] Chang, Yoo Kyung, Ioana Literat, Charlotte Price, Joseph I Eisman, Jonathan Gardner, Amy Chapman, and Azsaneé Truss. "News Literacy Education in a Polarised Political Climate: How Games Can Teach Youth to Spot Misinformation." Harvard Kennedy School Misinformation Review, 2020.

[90] BIK Portal. "Better Internet for Kids." Accessed September 14, 2021. https://www.betterinternetforkids.eu.

[91] "UNICEF: Make the Digital World Safer for Children – While Increasing Online Access to Benefit the Most Disadvantaged." Accessed September 14, 2021. https://www.unicef.org/press-releases/unicef-make-digital-world-safer-children-while-increasing-online-access-benefit-most.

[92] Calvo-Morata, Antonio, Cristina Alonso-Fernández, Manuel Freire, Iván Martínez-Ortiz, and Baltasar Fernández-Manjón. "Serious Games to Prevent and Detect Bullying and Cyberbullying: A Systematic Serious Games and Literature Review." *Computers & Education* 157 (November 1, 2020): 103958. https://doi.org/10.1016/j.compedu.2020.103958.

[93] Müller, Anna R., Mandy Röder, and Michael Fingerle. "Child Sexual Abuse Prevention Goes Online: Introducing 'Cool and Safe' and Its Effects." Computers & Education 78 (September 1, 2014): 60–65. https://doi.org/10.1016/j.compedu.2014.04.023.

[94] Scholes, Laura, Christian Jones, Colleen Stieler-Hunt, and Ben Rolfe. "Serious Games for Learning: Games-Based Child Sexual Abuse Prevention in Schools." International Journal of Inclusive Education 18, no. 9 (September 2, 2014): 934–56. https://doi.org/10.1080/13603116.2013.860195.

[95] Sanchawala, AadilMehdi J, Adhithya Arun, Rahul Sajnani, and Kavita Vemuri. "Unlocked: A Game On Human Trafficking." In 2020 IEEE Conference on Games (CoG), 415–22, 2020. https://doi.org/10.1109/CoG47356.2020.9231579.

[96] Roozenbeek, Jon, and Sander van der Linden. "Fake News Game Confers Psychological Resistance against Online Misinformation." Palgrave Communications 5, no. 1 (June 25, 2019): 1–10. https://doi.org/10.1057/s41599-019-0279-9.

[97] Katsaounidou, Anastasia, Lazaros Vrysis, Rigas Kotsakis, Charalampos Dimoulas, and Andreas Veglis. "MAthE the Game: A Serious Game for Education and Training in News Verification." Education Sciences 9, no. 2 (June 2019): 155. https://doi.org/10.3390/educsci9020155.

[98] UNICEF & Ministry for Foreign Affairs of Finland. "SomeBuddy." 2019. https://www.unicef.org/globalinsight/media/2086/file.

[99] UNICEF & Ministry for Foreign Affairs of Finland. "Milli Chatbot. Virtual Assistant at Mentalhub.fi." 2019. https://www.unicef.org/globalinsight/media/2081/file.

[100] Laleh, Naeimeh, Barbara Carminati, and Elena Ferrari. "Risk Assessment in Social Networks Based on User Anomalous Behaviours." IEEE Transactions on Dependable and Secure Computing 15, no. 2 (March 1, 2018): 295–308. https://doi.org/10.1109/TDSC.2016.2540637.

[101] Perez, Beatrice, Mirco Musolesi, and Gianluca Stringhini. "You Are Your Metadata: Identification and Obfuscation of Social Media Users Using Metadata Information." In Twelfth International AAAI Conference on Web and Social Media, 2018.

[102] Lee, Hee-Eun, Tatiana Ermakova, Vasilis Ververis, and Benjamin Fabian. "Detecting Child Sexual Abuse Material: A Comprehensive Survey." Forensic Science International: Digital Investigation 34 (September 1, 2020): 301022. https://doi.org/10.1016/j.fsidi.2020.301022.

[103] Pereira, Mayana, Rahul Dodhia, and Richard Brown. "Metadata-Based Detection of Child Sexual Abuse Material." ArXiv Preprint ArXiv:2010.02387, 2020.

[104] P. Anderson, Z. Zuo, L. Yang, and Y. Qu. "An Intelligent Online Grooming Detection System Using AI Technologies." In 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1–6, 2019. https://doi.org/10.1109/FUZZ-IEEE.2019.8858973.

[105] Rosa, H., N. Pereira, R. Ribeiro, P.C. Ferreira, J.P. Carvalho, S. Oliveira, L. Coheur, P. Paulino, A.M. Veiga Simão, and I. Trancoso. "Automatic Cyberbullying Detection: A Systematic Review." Computers in Human Behaviour 93 (April 1, 2019): 333–45. https://doi.org/10.1016/j.chb.2018.12.021.

[106] Fernández, Miriam, Alejandro Bellogín, and Iván Cantador. "Analysing the Effect of Recommendation Algorithms on the Amplification of Misinformation." ArXiv Preprint ArXiv:2103.14748, 2021.

Appendix A: Survey to minors

The questions asked in the survey of European minors are shown below:

1. *Order from MOST to LEAST the applications you use the most:*
   - *WhatsApp / Telegram*
   - *Instagram*
   - *TikTok*
   - *Snapchat*
   - *Twitter*
   - *Facebook*
   - *Online Video Games (with audio/text chats)*
   - *Dating Applications (Tinder, Grinder, etc.)*
   - *YouTube*
   - *Twitch*
2. *Are there any other applications that you use a lot that are not listed above?*
   - *[Open Answer]*
3. *Enter the number of hours you spend per day on the Internet for entertainment (social networking, gaming, etc.) on average on a weekday.*
   - ☐ *Less than 1 hour per day*
   - ☐ *Between 1 and 2 hours per day*
   - ☐ *Between 3 and 4 hours per day*
   - ☐ *More than 4 hours per day*
   - ☐ *Don't know / No answer*
4. *Enter the number of hours you spend per day on the Internet for entertainment (social networking, gaming, etc.) on average on a weekend day.*
   - ☐ *Less than 1 hour per day*
   - ☐ *Between 1 and 2 hours per day*
   - ☐ *Between 3 and 4 hours per day*
   - ☐ *More than 4 hours per day*
   - ☐ *Don't know / No answer*
5. *Please indicate which of the following devices you use the most. 1 being "I don't use it", and 5 being "I use it a lot".*
   - *Smartphone*
   - *Tablet*
   - *PC*
   - *Game Console (PlayStation, Xbox, Nintendo, etc.)*
   - *Smartwatch*
   - *Smart Speaker (Amazon Echo Alexa, Google Home, Apple HomePod, etc.)*

- *Smart Toys (i.e., toys that connect to smartphoes or the Internet)*

6. *Please indicate which of the following devices are your favourites. 1 being "I don't like it", and 5 being "I love it". If you do not use any of the devices you can indicate this by marking the option 0 "I don't use it".*

- *Smartphone*
- *Tablet*
- *PC*
- *Game Console (PlayStation, Xbox, Nintendo, etc.)*
- *Smartwatch*
- *Smart Speaker (Amazon Echo Alexa, Google Home, Apple HomePod, etc.)*
- *Smart Toys (i.e., toys that connect to smartphoes or the Internet)*

7. *From the point of view of cybersecurity and personal data privacy, which devices do you think are the most secure to use? Please rate the devices, with 1 being "Not secure", and 5 being "Very secure".*

- *Smartphone*
- *Tablet*
- *PC*
- *Game Console (PlayStation, Xbox, Nintendo, etc.)*
- *Smartwatch*
- *Smart Speaker (Amazon Echo Alexa, Google Home, Apple HomePod, etc.)*
- *Smart Toys (i.e., toys that connect to smartphoes or the Internet)*

8. *Do you know the Tor browser?*

- ☐ *Yes*
- ☐ *No*
- ☐ *Don’t know / No answer*

9. *If you answered 'Yes' to the above question, have you ever used it? What have you ever used it for?*

- ☐ *[Open Answer]*

10. *Rate the following Internet dangers according to the ones you are most concerned about. With 1 being "I am not concerned", and 5 being "I am very concerned".*

- *Fake News*
- *Dissemination of personal or intimate content without permission*
- *Online grooming (contact through the Internet by an adult with a minor for sexual purposes)*
- *Cyberbullying*
- *Identity theft or password theft*
- *Inappropriate content for young people*
- *Someone lying about who they are, their age and/or their gender*

11. *Have you ever been involved in a situation of racism that took place online? (e.g., insulting or laughing at a person because of their accent or skin color)*

- ☐ *It happened to me*
- ☐ *I did it*
- ☐ *I know someone who happened to him/her*
- ☐ *I know someone who did it*
- ☐ *Don't know / No answer*

12. *And with any online situation involving LGTBIphobia? (e.g., insulting or making fun of someone because of their sexual orientation, a boy for being effeminate or more masculine)*

- ☐ *It happened to me*
- ☐ *I did it*
- ☐ *I know someone who happened to him/her*
- ☐ *I know someone who did it*
- ▪ *Don't know / No answer*

13. *Have you found any sexist situations online? (e.g., sexist jokes, or inappropriate comments about girls' physical appearance)*

- ☐ *It happened to me*
- ☐ *I did it*
- ☐ *I know someone who happened to him/her*
- ☐ *I know someone who did it*
- ▪ *Don't know / No answer*

14. *What about any display of violent or demeaning content online? (For example, any video showing a person being seriously injured or animal abuse)*

- ☐ *It happened to me*
- ☐ *I did it*
- ☐ *I know someone who happened to him/her*
- ▪ *I know someone who did it*
- ▪ *Don't know / No answer*

15. *Think about whether, in the last year, you remember any situation that involved receiving repeated insults through social networks.*

- ☐ *It happened to me*
- ☐ *I did it*
- ☐ *I know someone who happened to him/her*
- ▪ *I know someone who did it*
- ▪ *Don't know / No answer*

16. *What about posting an embarrassing video, photo or meme of you/another person, without permission?*

- ☐ *It happened to me*
- ☐ *I did it*

- ☐ *I know someone who happened to him/her*
- *I know someone who did it*
- *Don't know / No answer*

17. *Did someone you know break into your/ someone else's account without permission?*

- ☐ *It happened to me*
- ☐ *I did it*
- ☐ *I know someone who happened to him/her*
- *I know someone who did it*
- *Don't know / No answer*

18. *Did someone try to isolate you/that person from a group of friends or class?*

- ☐ *It happened to me*
- ☐ *I did it*
- ☐ *I know someone who happened to him/her*
- *I know someone who did it*
- *Don't know / No answer*

19. *Can you think of a situation where someone sent unwanted sexual pictures to a person of a similar age?*

- ☐ *It happened to me*
- ☐ *I did it*
- ☐ *I know someone who happened to him/her*
- *I know someone who did it*
- *Don't know / No answer*

20. *Have you experienced or witnessed any other situation related to cyber-bullying / cyber-harassment that you have not previously indicated?*

- ☐ *[Open Answer]*

21. *Have you received a friend request from an unknown adult (over 18)?*

- ☐ *It happened to me*
- ☐ *I know someone who happened to him/her*
- *Don't know / No answer*

22. *Have you received messages or photos from an unknown adult (over 18)?*

- ☐ *It happened to me*
- ☐ *I know someone who happened to him/her*
- *Don't know / No answer*

23. *Have you had conversations with an unknown adult (over 18)?*

- ☐ *It happened to me*
- ☐ *I know someone who happened to him/her*
- *Don't know / No answer*

24. *Have you shared news/stories, and later found out they were false?*

- ☐ *It happened to me*

- ☐ *I know someone who happened to him/her*
- *Don't know / No answer*

25. *Have you shared news/stories, knowing they were false?*

- ☐ *It happened to me*
- ☐ *I know someone who happened to him/her*
- *Don't know / No answer*

26. *When it comes to sharing news/stories on the Internet, which of the following options are most important to you? Answer each option, with 1 being "I don't care at all", and 5 being "I care a lot".*

- *My followers are interested in*
- *I trust the person I received it from*
- *The information comes from a newspaper, media or a known author*
- *I have checked with other sources that the information is correct*
- *The information I am going to share is funny to me*
- *I prefer to share news/stories in video or photo format*
- *I prefer to share news/stories in text format*

27. *Regarding risks on the Internet (indicate None / A little / Quite / A lot)*

- *At home they have talked to me about these dangers*
- *At home they monitor what I do on the Internet*
- *I have been given information at school (teachers or outside experts).*
- *I think it would be useful to have more information*

28. *If you think it would be useful to have more information about risks on the Internet, what specific topics would you like to know more about?*

- ☐ *[Open Answer]*

29. *Socio-demographic*

- *Age:*
- *Gender: Male / Female / Non-binary / Prefer not to say*
- *Sexual orientation: Heterosexual / Homosexual / Bisexual // Other / Not clear / Prefer not to say*
- *Country of residence:*
- *City:*
- *Mother's country of birth:*
- *Father's country of birth:*