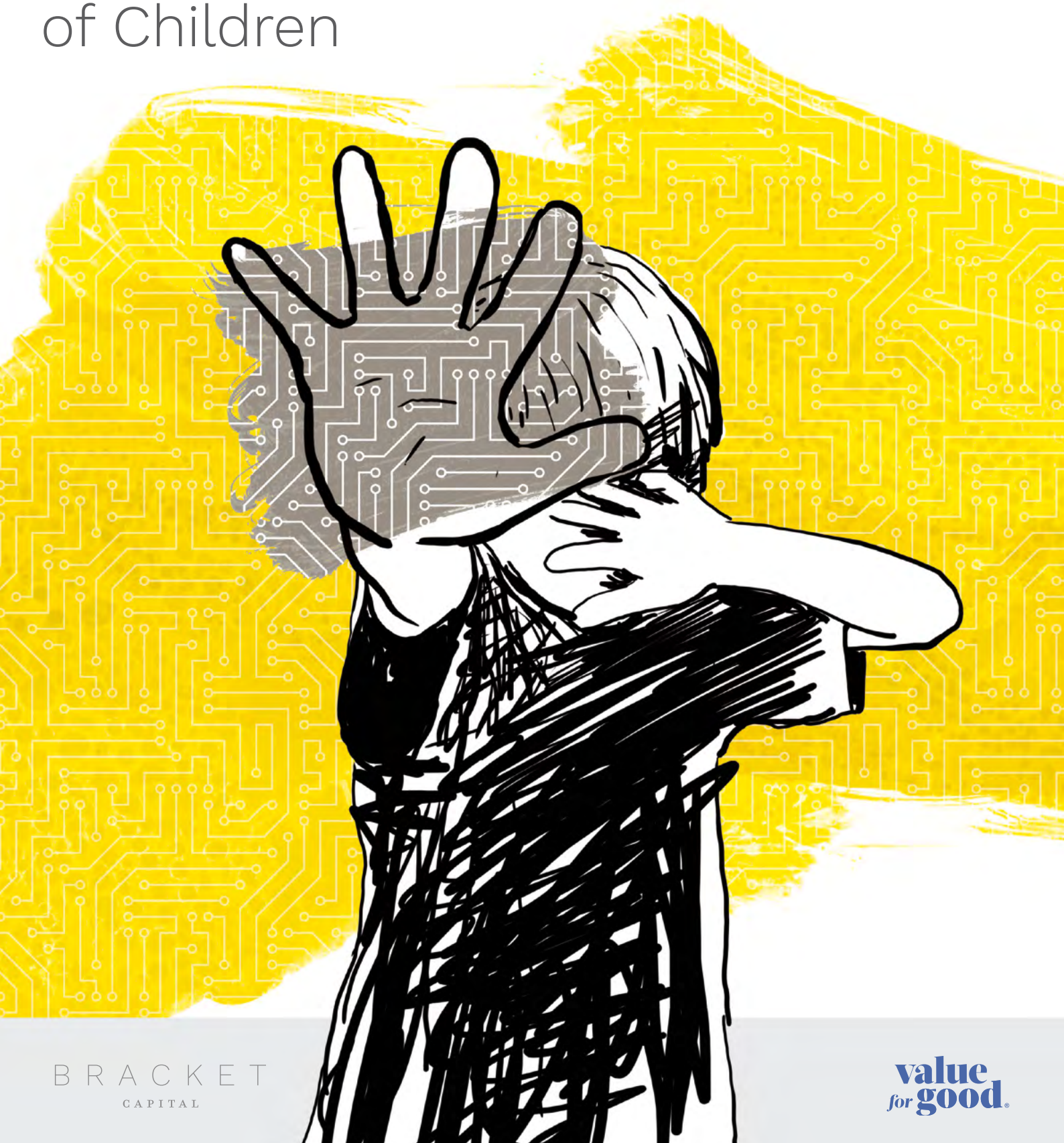
BRACKET
Foundation

Artificial Intelligence

Combating Online Sexual Abuse of Children



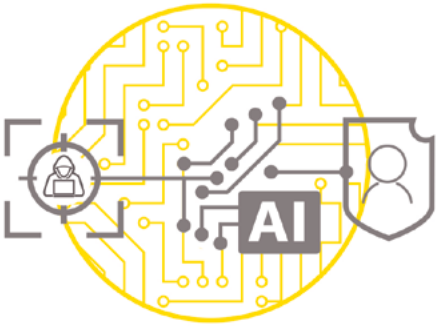
BRACKET
CAPITAL

value for good

The digital era has revolutionized our lives, but it also fueled an **unprecedented surge in online sexual abuse of children**

The scale, variety and severity of online child sexual abuse **challenges prevention, detection and prosecution** efforts

AI has significant potential to help fight the increasing scale of online sexual abuse of children through its ability to process massive volumes of data and learn basic human tasks



A mapping of the global landscape of tools revealed approximately 50 AI-based technologies which are beginning to **improve prevention, detection and prosecution** efforts

More must be done with AI to safeguard children from online sexual abuse: by (1) upgrading traditional solutions with AI, (2) developing more AI solutions for **prevention** and **prosecution** and (3) expanding AI solutions to new geographies whilst tackling new forms of abuse

Action is required by multiple stakeholders to overcome the key barriers preventing the wider adoption of AI to (1) protect children from becoming victims, (2) disrupt abuse when it happens and (3) effectively bring perpetrators to justice

Contents

Executive Summary

The rapid growth of digital technology has revolutionized our lives, transforming the way we connect and communicate. Internet access, mobile devices and social media are now ubiquitous, especially among children. Of the 4.5 billion people with access to the internet today, 1 in 3 are under the age of 18, and they are often unsupervised.

The dark side of this development has been an **unprecedented surge in online sexual abuse of children**. Not only are children increasingly exposed to abusive behavior online, but today 1 in 5 children are sexually solicited while online. Perpetrators profit from technological advances such as easy sharing of large files, fast live streaming of videos and strong encryption software. Indeed, nearly every facet of online sexual abuse has been made easier by technology, leaving children of all ages more vulnerable than ever online. This **"digital era of abuse"** gives rise to three alarming trends that make the problem especially difficult to combat:

1. **Traditional forms of child sexual abuse** have increased materially resulting in a record number of both children prostitution and distribution of child sexual abuse material (CSAM) online.

2. **New forms of digital and mobile abuse** – online grooming and live-streamed abuse, have emerged rapidly and at large scale.

3. The **severity of abuse and the damage to victims has increased** – pedophiles, among others, encourage each other through online platforms and victims are unable to stop the repeated sharing of their content.

These three alarming trends come together to create a problem that is global, fast-growing and extremely difficult to combat. This **problem** is leaving victims, parents, institutions and technology providers struggling with its **prevention, detection and prosecution**.

This is where **Artificial Intelligence (AI) can be leveraged**. AI can conduct analysis, provide decision recommendations and carry out actions at a scale, speed and depth of detail not possible for human analysts. Some examples include:

- Applying machine learning to help parents monitor their children's activity online to **prevent** abuse from happening

- Image classification algorithms which automatically identify CSAM to **detect** abuse after it has occurred without requiring human reviewers

- Chatbots which carry out text-message conversations with buyers and providers of sexual services to gain additional information to **prosecute** perpetrators faster

These and other AI tools are still being piloted in the fight against child sexual abuse online, but show significant potential in **preventing, detecting and prosecuting crimes** more efficiently and effectively. Analysis of the emerging technology shows that by building on a strong base of proven AI capabilities, there are three key ways AI can support the fight:

1. **Upgrade** existing manual and analytics solutions with the latest AI capabilities

2. **Shift** development of new AI tools from **detection** to **prevention and prosecution**

3. **Expand** the reach of AI solutions to new geographies and new forms of abuse

Bringing the full strength of AI technology to the fight against online sexual abuse of children cannot be achieved without the full participation and concerted efforts from key stakeholders. The path forward will require the following:

1. **Share existing knowledge and increase collaboration among stakeholders**
   Technology players need to bring in their AI know-how; law enforcement agencies their intelligence on perpetrators and criminal activity trends; and civil society actors their deep understanding of issues faced by victims to ensure solutions respect the rights of children.

2. **Establish new forms of collaboration across sectors and borders**
   The different stakeholder groups should collaborate across borders and sectors with an unprecedented determination and scale to outperform the progress made by massive global communities of perpetrators online.

3. **Redefine legal frameworks and cooperation agreements enabling secure use and sharing of data**
   New and harmonized legal frameworks are essential to facilitate cooperation between multiple stakeholders and to enable secure data transfer that respect data privacy regulations, such as GDPR.

4. **Allocate more resources to develop and expand AI solutions proactively**
   Public authorities and private actors need to closely collaborate and jointly invest in tools that improve detection, but more importantly in tools that can disrupt and eliminate abuse at its source.

5. **Increase awareness and understanding of the severity of the problem and its many forms**
   More awareness among the general public on the magnitude of online sexual abuse of children and available solutions is required to increase the understanding of its alarming new forms and facets.

6. **Invest in the development of enhanced digital skills for both law enforcement and civil society**
   Build knowledge, understanding and trust in AI as a credible and scalable solution to online child sexual abuse. Law enforcement agencies need sufficient expertise in data science as well as a better understanding of the tools currently used by perpetrators so they can win this fight. **Non-profit actors and parents need to better understand the risks currently incurred by children in order to make better decisions to keep them safe online**.

AI is one of many tools to combat the online sexual abuse of children. **Traditional efforts will remain important** and can be **enhanced by AI**. There is little doubt that online connectivity is leading to a dramatic increase in online sexual abuse. As a society, we must strive to make the internet safer for children. Technology may be the source of the problem but it may also be the solution we aspire for.

1 in 3
internet users
is under the
age of 18

I. The Magnitude of the Problem

The rapid growth of digital technology has revolutionized our lives, transforming the way we connect and communicate. Internet access, mobile devices, social media and messaging apps are now ubiquitous, especially among children. Of the 4.5 billion people with access to the internet today, 1 in 3 are under the age of 18 and often unsupervised.[1]

The dark side of this development has been an unprecedented surge in online sexual abuse of children. The global spread of new technology has digitalized all walks of life – often with little awareness of the risks new technologies pose to children. Indeed, nearly every facet of online sexual abuse has been made easier by technology, leaving children of all ages more vulnerable than ever online.

The children that are especially vulnerable to becoming a victim of online sexual abuse often also experience challenges in life offline. These challenges often include severe living conditions, such as homelessness; experiences of physical or sexual abuse offline; growing up in reconstituted families; high levels of conflict with parents; and mental health difficulties. Demographic and personal behavioral factors can also play a role[2]. Being a girl, aged between 13 and 15 years or of black ethnicity increases the risk of being exposed to sexual harassment, solicitation and grooming online.[3] Similarly, children who use the internet frequently, especially without parental monitoring, are more likely to be targets of online sexual abuse. A higher level of education and knowledge of online sexual abuse can reduce these risks.[4]

This "digital era of abuse" gives rise to three alarming trends that make the problem especially difficult to combat: (1) the explosion in scale of traditional forms of child sexual abuse; (2) the emergence of new forms of digital and mobile abuse; (3) and the increased severity of abuse and damage to victims.

Alarming Trend #1: Surge in traditional forms of sexual abuse of children

Traditional forms of abuse that predate the internet have scaled up exponentially as perpetrators embrace new technologies and online channels. These include child pornography (referred to as CSAM)[5] and child sex trafficking. Abuse has escalated from dozens of low-quality images to sophisticated cyberlockers hosting HD videos; from cryptic advertisements on the surface web to detailed reviews of individual children on dark web forums.

Online child sexual abuse is scaling with strong digital infrastructure and low-cost hosting. **95% of the world's CSAM is hosted in Europe and North America,** with the Netherlands alone hosting 47%.[6] The problem is also spreading rapidly to emerging markets as they gain full access to modern technology and connectivity. This increasing globalization of online child sexual abuse is cause for alarm and **action**.

There has been an exponential growth in CSAM available online. Over the past decade the volume of images and videos of suspected child sexual abuse reported to the US National Center for Missing & Exploited Children (NCMEC) has exploded from 450,000 files in 2004 to more than 45 million files in 2018. At the same time, the number of reports of URLs containing CSAM has increased from only 3,000 in 1998 to 18.4 million today.[7]

1 Sonia Livingstone, John Carr, Jasmina Byrne. "One in Three: Internet Governance and Children's Rights". UNICEF Office of Research – Innocenti. January 2016.

2 Aiman El Asam & Adrienne Katz. "Vulnerable Young People and Their Experience of Online Risks". Journal Human-Computer Interaction. February 2018

3 Kimberly J. Mitchell et al. "Online Requests for Sexual Pictures from Youth: Risk Factors and Incident Characteristics". Journal of Adolescent Health. August 2007

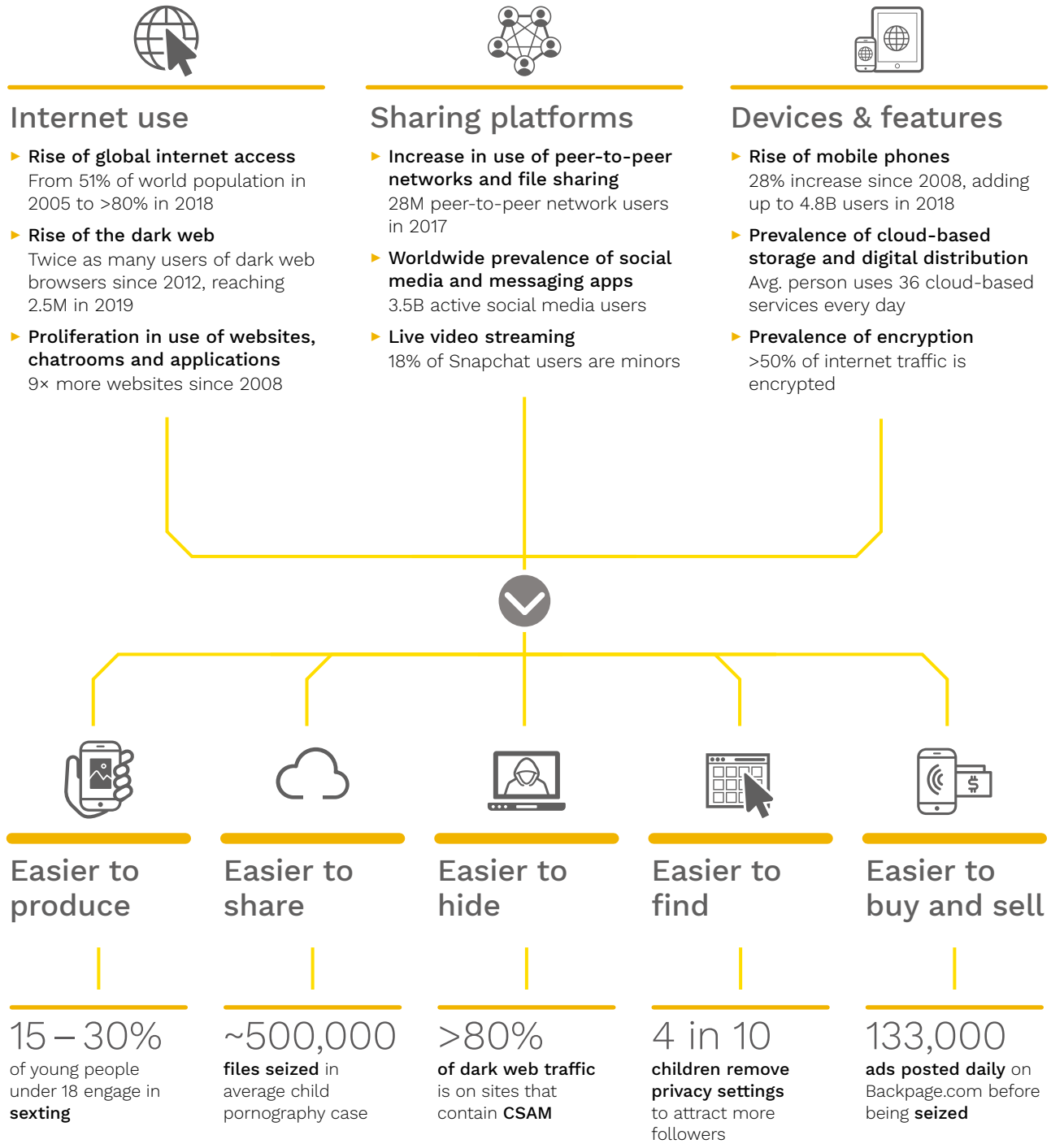
4 Mare Ainsaar & Larse Lööf. "Online behaviour related to child sexual abuse – Literature Report".

5 CSAM (Child Sexual Abuse Material) is the preferred term as it connotes the abusive nature of images; pornography on the other hand implies child's consent.

6 The Internet Watch Foundation. "IWF Annual Report 2018: Once Upon a Year". April 2019.

7 Julie Cordua. "A Bold Goal: Eliminating Child Sexual Abuse from the Internet". Thorn. April 2019.

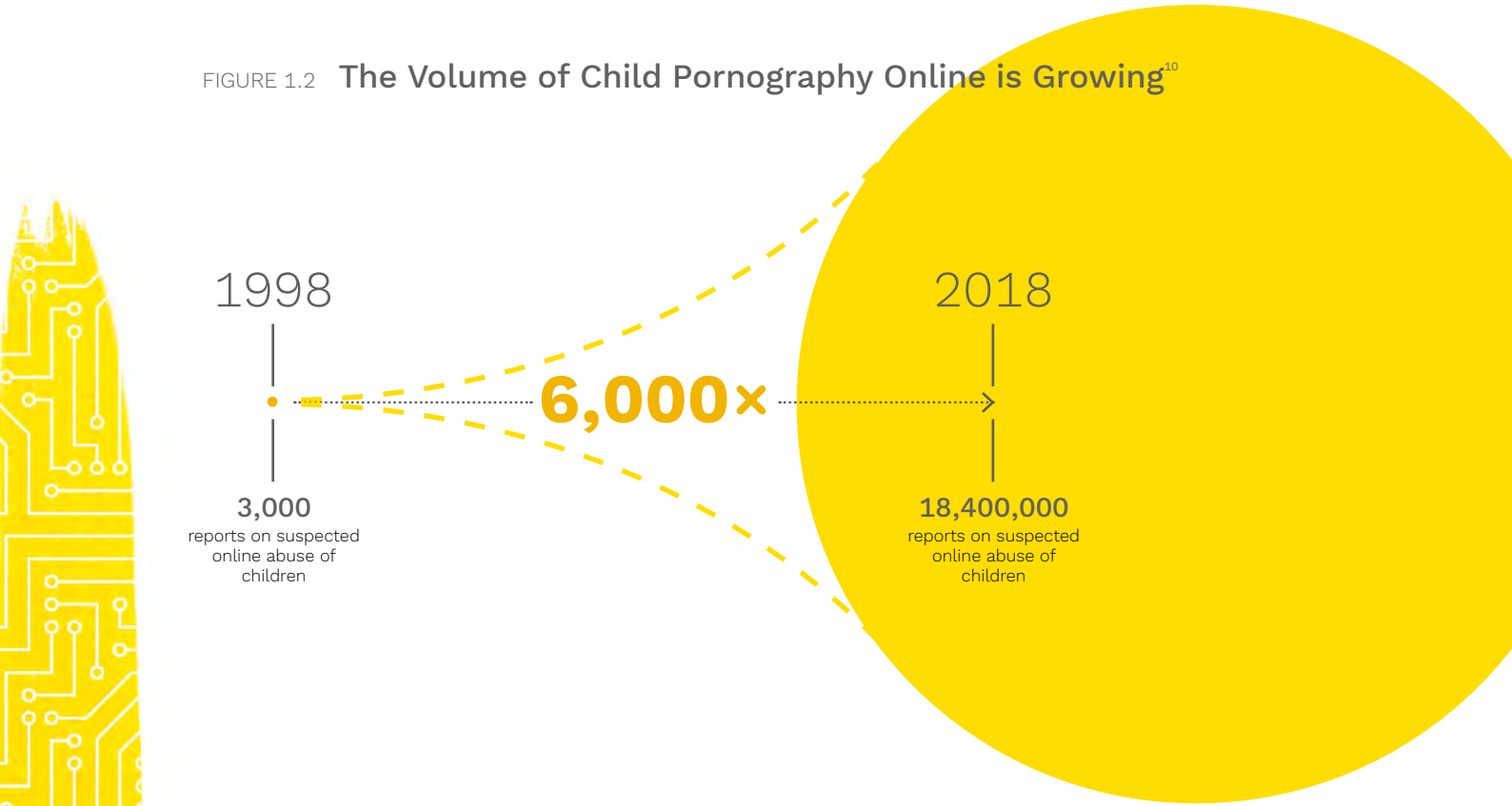
FIGURE 1.1 **Advances in technology make it easier to abuse children online**[8,9]



8 Sources for upper part of figure: ITU. "ITU releases 2018 global and regional ICT estimates". December 2018; The Tor Project. "User Metrics". August 2019; Internet Live Stats. "Total Number of Websites". August 2019; Cisco. "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update". February 2019; Sandvine. "The Global Internet Phenomena Report". October 2018; Tecxipio. "Tecxipio Magazine: File Sharing"; We are social. "Digital in 2019". January 2019; Statista. "Distribution of Snapchat users worldwide as of July 2019, by age and gender". July 2019; Statista. "Number of mobile phone users worldwide from 2015 to 2020". November 2016; Petapixel. "The Importance of Cameras in the Smartphone War". February 2015; Dr Garrick Hileman, Michael Rauchs. "Global Cryptocurrency Benchmarking Study". 2017; Techjury. "Clould Computing Statistics 2019". March 2019.

9 Sources for bottom part of figure: Sheri Medigan, Anh Ly, Christina L. Bash. "Prevalence of Multiple Forms of Sexting Behavior Among Youth – A Systematic Review and Meta-analysis". April 2018; NetClean. "NetClean Report 2017", 2018; Dr Gareth Owenson, Dr Nick Savage. "Empirical analysis of Tor hidden services". May 2016; Internet Matters Ltd. "Infographic revealing kids' use of social media survey stats"; Marinus Analytics. "Marinus Analytics finds sex trafficking surging online after Backpage.com shutdown". November 2018.

FIGURE 1.2 **The Volume of Child Pornography Online is Growing**[10]



The surge in CSAM is driven by the advent of peer-to-peer file sharing services and low-cost digital distribution, which enable anonymous access to massive collections including millions of images. The trade in CSAM is largely a non-commercial activity where large volumes of material are shared among like-minded perpetrators at no cost.[11] The dark web is where communities of abusers meet and share links and encryption keys to unlock repositories of CSAM, which are still in large part hosted on the surface web. Facebook, for example, reports removing at least 8 million abusive images of children each quarter.[12]

The rapid growth of digital platforms has facilitated child sex trafficking and transformed it into a **highly lucrative industry.** Just as eBay and Amazon revolutionized the process of finding, reviewing and ordering products online, traffickers have used platforms to streamline recruiting, advertising and the sale of children for sex. These lower costs have led to a spike in volume and profitability of the crime; although sex trafficking represents only 20% of global human trafficking victims, it makes up 66% of the profits, with victims generating a return on investment between 100% and 1000% for traffickers.[13] Globally more than 1 million children are trafficked for sex.[14] These children are hidden in plain sight in floods of online sex advertisements: for example, in the US 150,000 new escort advertisements are posted every day.[15] Traffickers place advertisements using keywords and misspellings known to perpetrators looking to have sex with children. While most buyers have low technological sophistication, specialized online forums and hobby boards exist where perpetrators shop for sex with minors and screen reviews and ratings of individual children.[16] 75% of sex trafficking survivors today report they were being advertised online, compared to only 38% in 2004. The number of daily buyers has risen dramatically as traffickers use modern communications tools to engage victims and buyers on popular sites; 1 in 4 sex trafficking survivors advertised online reported more than 10 buyers per day, compared to just 1 in 7 for those advertised offline on the street.[17] Commercial escort websites are occasionally shut down for sex trafficking (such as Backpage in 2018); however, closures tend to drive activity to new sites and the dark web.

10 Based on a number of tips reported to the US-based National Center for Missing and Exploited Children (NCMEC)

11 ECPAT International. "Briefing Note to Committee on the Rights of the Child". September 2014.

12 Reuters. "Facebook removes 8.7 million sexual photos of kids in last three months". October 2018.

13 Human Rights First. "Fact Sheet: Human Trafficking by the Numbers". September 2017.

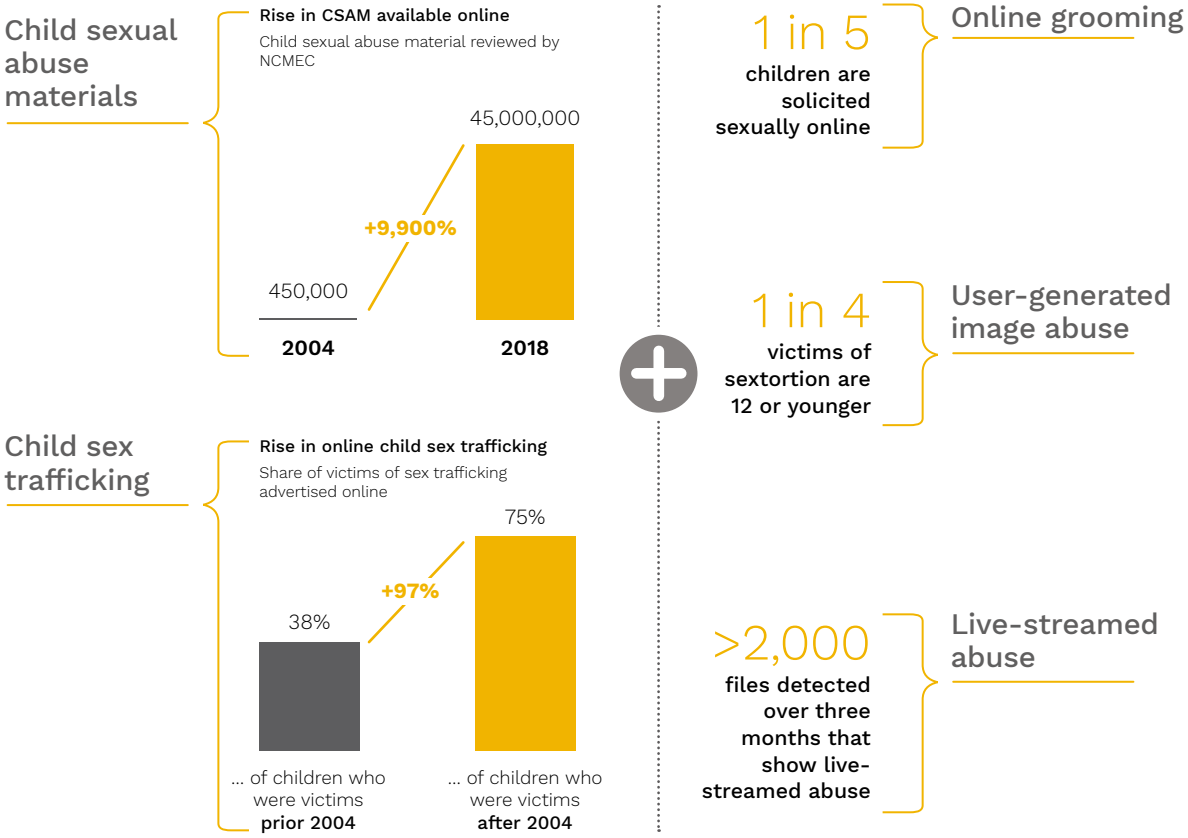
14 International Labour Organization. "Global Estimates of Modern Slavery". Alliance 8.7.2017.

15 Thorn. "Child Sex Trafficking Statistics".

16 Rob Spectre, Marian Hatcher. "National Johns Suppression Initiative". childsafe.ai, Cook County Sheriff. 2018.

17 Thorn, Dr. Vanessa Bouche, Texas Christian University. "Survivor Insights: The Role of Technology in Domestic Minor Sex Trafficking". January 2018.

FIGURE 1.3 **As traditional forms of abuse grow, new digital forms emerge**[18, 19]



Alarming Trend #2: Emergence of new forms of digital and mobile abuse

Technology has created new mechanisms for all-digital abuse, including online grooming, user-generated image abuse and live-streamed abuse, which are impacting millions of children who were previously safe from online abuse.[20]

Online grooming refers to when an adult wrongfully gains the trust of a child online and then convinces the child to commit sexual acts. Popular social media platforms, gaming sites and child-friendly websites have become breeding grounds for online grooming. These platforms provide access to freely shared personal information on victims that abusers use to find, stalk and bully victims online. As many as 1 in 5 children are solicited sexually while on the internet.[21] Once they gain online access to children, perpetrators use a mix of persuasion tactics to lure children, such as "catfishing" – impersonating other young people to gain their trust – and get them to share sexual images of themselves or provoke other sexual behavior. Children are uniquely vulnerable to grooming with unsupervised use of the internet, smartphones and webcams, which they use to share images and communicate with strangers.

User-generated image abuse is when images of children that are innocently produced are used to bully and exploit children sexually online. The images are either generated by the victims themselves or created by perpetrators with digital tools, such as clipping images from innocent YouTube videos or producing "deep fake" pornography where a child's face is digitally pasted onto existing CSAM.

18 Thorn. "Technology has made it easier to harm kids"; Thorn, Dr. Vanessa Bouche, Texas Christian University. "Survivor Insights: The Role of Technology in Domestic Minor Sex Trafficking". January 2018.

19 Anja Schulz et al. "Online Sexual Solicitation of Minors: How Often and between Whom Does It Occur". Journal of Research in Crime and Delinquency. 2016; Janis Wolak & David Finkelhor. "Sextortion: Findings from a survey of 1,631 victims". June 2016; Internet Watch Foundation. "Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed Child Sexual Abuse". May 2018.

20 Darkness to Light. "Child Sexual Abuse Statistics Report". 2015.

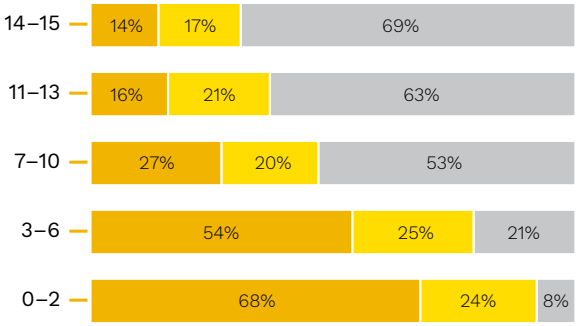
21 Kimberly Mitchell et al. "Risk Factors for and Impact of Online Sexual Solicitation of Youth". Journal of the American Medical Association. 2018.

This emerging type of abuse can also take the form of "sextortion", where the child is blackmailed using self-generated sexting images to extort sexual favors, under threat of sharing the images on social media or with family members. Technology has driven the proliferation and loss of control of self-generated content. Up to 88% of self-generated, sexually explicit online content has been taken from its original location and uploaded elsewhere. "Sexting" behavior makes adolescents especially vulnerable to abuse. Around 15-40% of young people engage in sexting, using smartphones, messaging apps and lives-streaming technology to explore their sexuality in an increasingly risky online environment.[22]

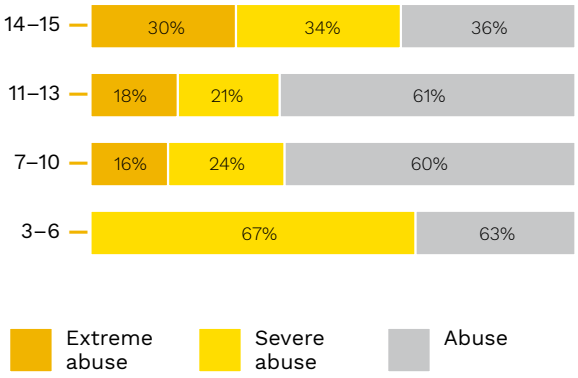
Live-streamed abuse has emerged out of advances in streaming and mobile camera technology. With just a camera-enabled device and internet connection, it is possible to facilitate a live sex show for paying customers anywhere in the world. This crime, typically facilitated by a family member or known adult in the home of the victim, transcends national borders, with wealthy perpetrators concentrated in high-income countries paying large sums for short, on-demand abuse sessions taking place in lower-income countries. The phenomenon is well-documented in Southeast Asia but is spreading to other regions.[23] It is very difficult to accurately measure the growing magnitude of the problem given the inability to **detect** the faint and temporary digital traces that live-streamed abuse leaves behind.

FIGURE 1.4 **Abuse is shifting towards digital and more severe forms**

URLs hosting child sexual abuse content by age of children and severity of abuse[24]



Live-streamed child sexual abuse by age of child and severity of abuse[25, 26]



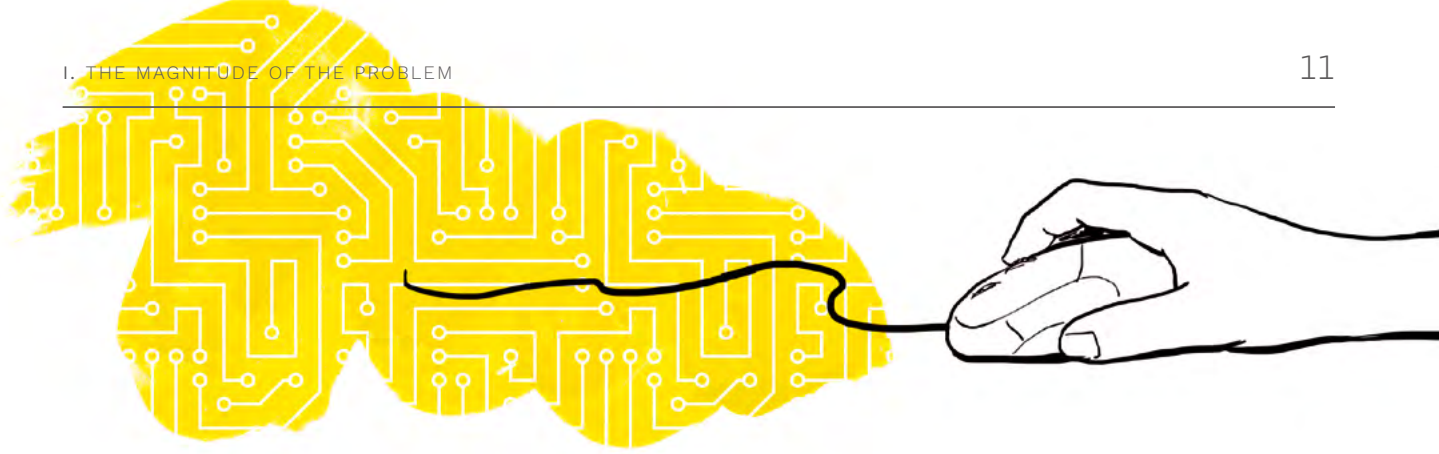
22 Jessica Ringrose et al. "A qualitative study of children, young people and 'sexting': a report prepared for the NSPCC". National Society for the Prevention of Cruelty to Children. 2012.

23 Deanna Davy. "The Sexual Exploitation of Children in Southeast Asia". ECPAT International. September 2017.

24 The Internet Watch Foundation. "IWF Annual Report 2018: Once Upon a Year". April 2019.

25 Internet Watch Foundation. "Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed Child Sexual Abuse". May 2018.

26 As defined by IWF: Extreme Abuse (Category A): Penetrative sexual activity, incl. rape and sexual torture; Severe Abuse (Category B): non-penetrative sexual activity, Abuse (Category C): indecent images/videos not in Category A or B.



Alarming Trend #3: Increased severity and damage of abuse

Perpetrators have new, technologically sophisticated ways of abusing children online without being detected. Digital technology has empowered the worst child sexual abusers to join together in virtual global communities. These networks of abuse can be highly prolific and tend to be technologically sophisticated. They "trade" large volumes of sexual abuse imagery online and create communities to share tips and tricks for finding new victims and avoiding law enforcement **detection**. The scale of these communities is alarming – the online forum PlayPen alone had 150,000 members before it was taken down.

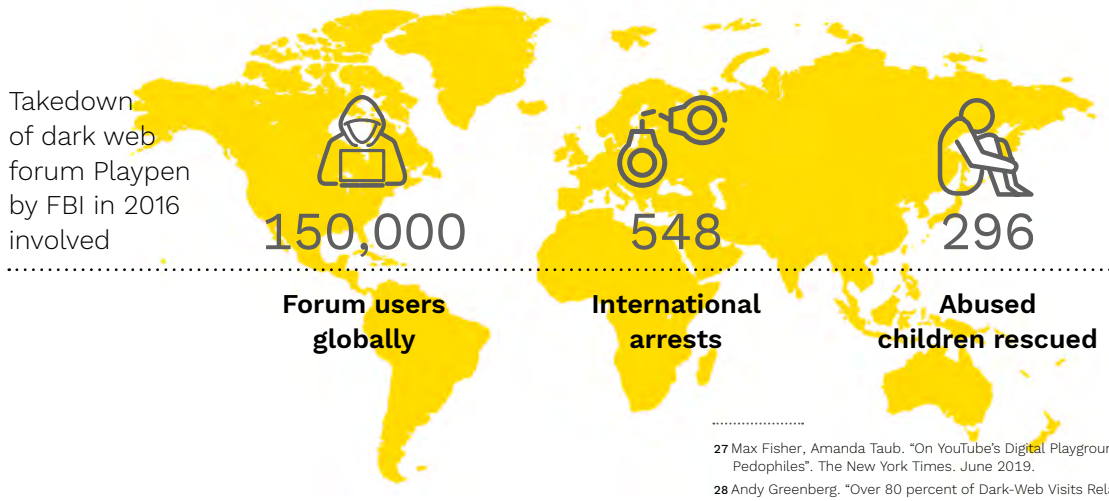
Pedophile communities are leveraging mainstream online platforms to connect with each other and abuse children in plain sight. For example, a NEW YORK TIMES article reported that YouTube's recommendation algorithms were helping pedophiles be directed to innocently shared videos of children. In the comments sections, communities of pedophiles could be found making sexually explicit remarks about the videos and reacting to each others' posts.[27]

The dark web has become a powerful tool for more serious perpetrators, with more than 80% of dark web traffic being generated by visits to sites with CSAM.[28] With the anonymity of encrypted channels and mobile devices, perpetrators now operate with less effort and lower risk of **detection**. The most sophisticated perpetrators carefully cover their digital tracks online.

Victims are now more vulnerable than ever to online sexual abuse, across all ages and online environments. While children of all ages are sexually abused online, there is a clear trend toward younger children and more extreme acts of sexual violence. The Internet Watch Foundation (IWF) found that 39% of CSAM online is of victims under the age of 10, and 43% depicts acts of extreme sexual violence.[29]

FIGURE 1.5 Global online communities of perpetrators are growing[30]

THE CASE OF PLAYPEN FORUM



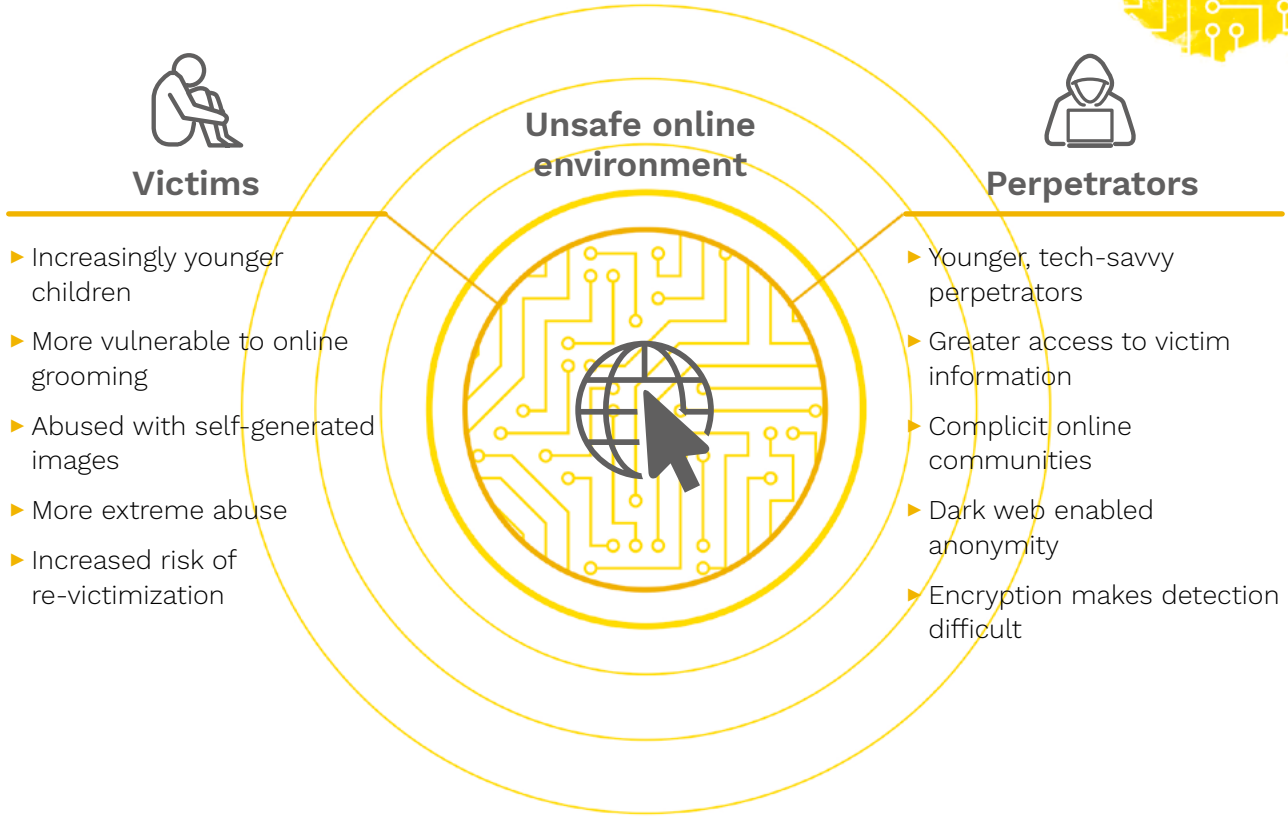
27 Max Fisher, Amanda Taub. "On YouTube's Digital Playground, an Open Gate for Pedophiles". The New York Times. June 2019.

28 Andy Greenberg. "Over 80 percent of Dark-Web Visits Relate to Pedophilia, Study Finds". WIRED. December 2014.

29 The Internet Watch Foundation. "IWF Annual Report 2018: Once Upon a Year". April 2019.

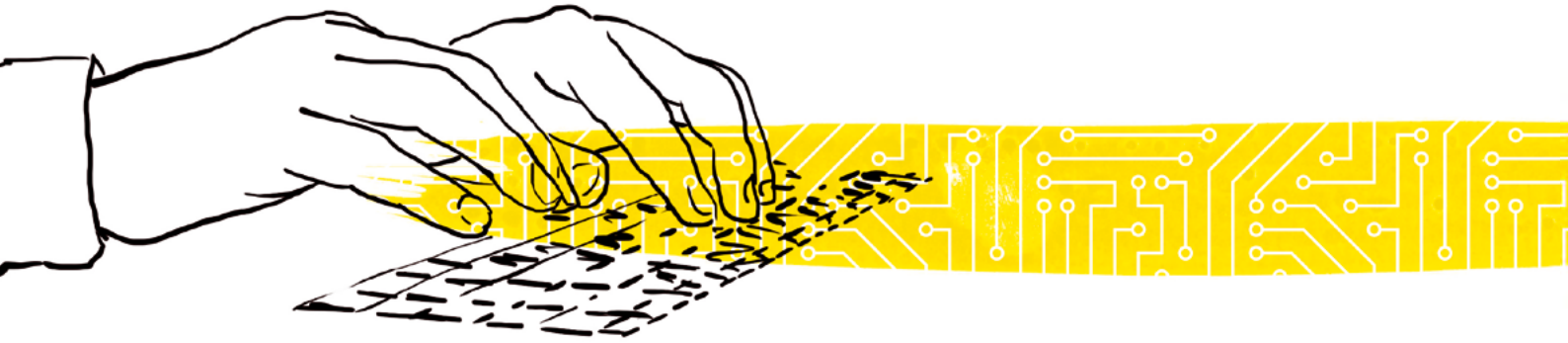
30 Elie Bursztein. "Rethinking the Detection of Child Sexual Abuse Imagery on the Internet". Google, in collaboration with NCMEC and Thorn. May 2019.

FIGURE 1.6 **Digital era intensifies victimization and abuse**



The younger the child, the more extreme the form of sexual abuse tends to be. An IWF study of live-streamed abuse revealed that 63% of images of 0- to 2-year-olds involved the worst forms of abuse, versus 20% for 11- to 13-year-olds and 7% for 16- to 17-year-olds.[31] Meanwhile, the 11- to 13-year-old age group has seen a massive spike in abuse arising from self-generated images.

New digital forms of sexual abuse have a **devastating impact on children,** as various forms of online abuse are often layered and intertwined. Children are left feeling trapped in a cycle of continuous abuse across devices and platforms, unable to escape their online abusers. The repeated sharing of images and videos revictimizes children, intensifying feelings of shame and powerlessness that cause long-term psychological damage.[32]



31 The Internet Watch Foundation. "Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed Child Sexual Abuse". May 2018.

32 Canada Centre for Child Protection. "Survivors' Survey: Full Report 2017". Protectchildren.ca. 2017.

These three alarming trends create challenges for all stakeholders, making it difficult to **prevent, detect and prosecute** online sexual abuse of children.

Failure to **prevent**

Parents, caregivers and policymakers are are **failing to come to terms with the severity** of the issue. On the one hand, children are gaining an increasingly sophisticated understanding of the internet and mobile phone technologies and are sharing more potentially compromising images and videos online. On the other, parents, caregivers and policymakers often have more rudimentary digital skills and lack a basic awareness of the risks their children are exposed to.[33] The gap between children and those charged with their safety can be particularly significant in regions where internet penetration has grown exponentially in recent years such as in Africa. Children are typically left to their devices, internet and social media in an unsupervised, unmonitored way.

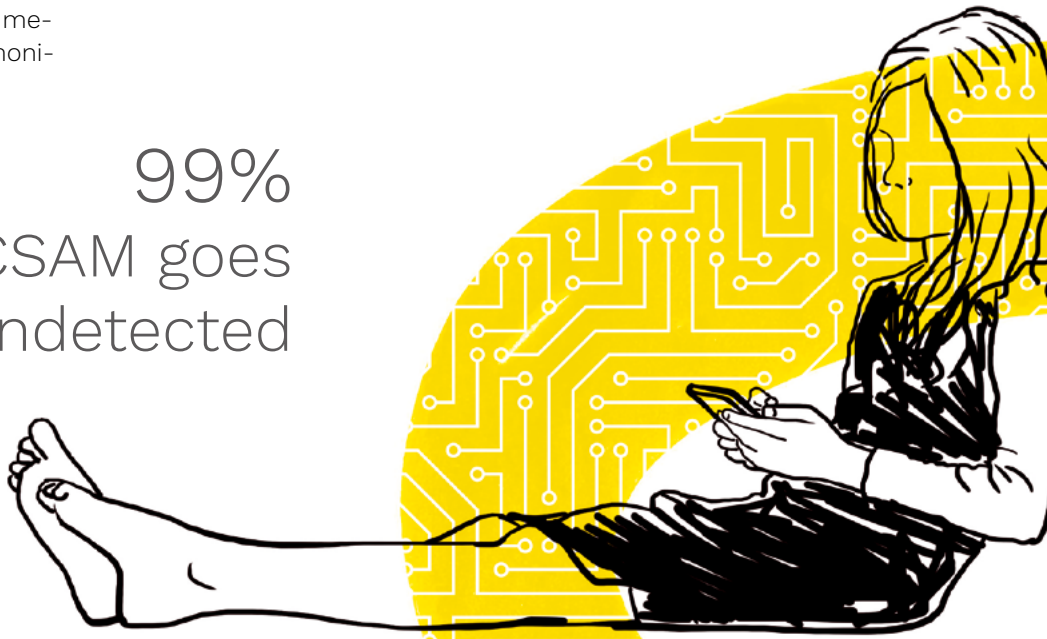
Failure to **detect**

Internet and technology companies are finding it challenging to **detect and identify abuse** across the web due to the growing volume of abusive activity online. According to estimates, on average less than 1% of CSAM uploaded on the internet gets identified for potential removal, leaving 99% of CSAM undetected. Image hosting sites and social media sites are breeding grounds for abuse, and given the proliferation of sites, it remains challenging to spot the abusive content and remove it at its source.

Failure to **prosecute**

Law enforcement and NGOs are facing difficulties in pursuing perpetrators and bringing them to justice. They are challenged by the massive volumes of data involved in an average CSAM or global sex trafficking case and by the exponential growth of CSAM and sex trafficking reports received through hotlines. This volume has outstripped the capability of human analysts to process it, leaving law enforcement overwhelmed with significant backlogs of potentially criminal material – without the resources or technical capacity to do so. As a result, the vast majority of perpetrators are never **prosecuted,** and globally only 1% of sex trafficking victims are rescued.[34]

99%
of CSAM goes
undetected



33 ECPAT. "Briefing Note to Committee on the Rights of the Child". September 2014.

34 Simone Monasebian. United Nations Office on Drugs and Crime. July 2016.

II. How AI Can Help

The challenges of **preventing, detecting** and **prosecuting** online child sexual abuse – particularly given its unprecedented global scale and complexity – require technological solutions. This is where AI can bring its strengths to the fight. AI can draw conclusions, solve problems or take actions by analyzing options and reasoning without the need for hard-coded instructions for each and every scenario. It builds its intelligence by learning from historical data, using statistical analysis.

AI can conduct analysis and provide decision recommendations at a **scale, speed and depth of detail not possible for human analysts.** While analytics technology already alleviates the "human workload" by aggregating and structuring data for human consumption, it still requires significant human input and interpretation to take action. The recommendations and predictions offered by AI improve on those offered by traditional analytics technology because they use self-learning algorithms to improve through "experience" – learning more complex tasks by adding more data and computing power to analytics models.

FIGURE 2.1 **Evolution from data analytics to AI**

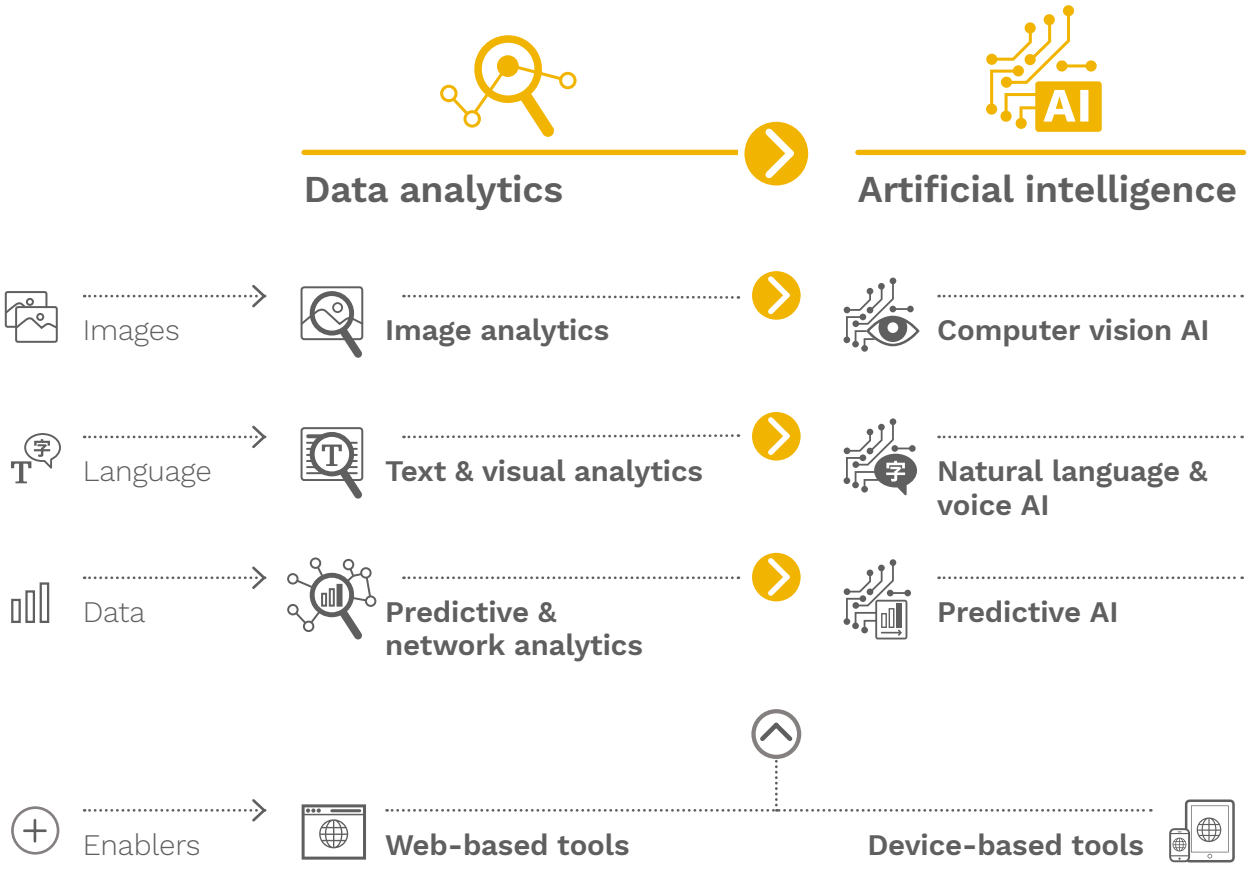


FIGURE 2.2 **Evolving image tools**

Image hashing technology detects **known CSAM online**

1. Collect potential child abuse image by working with law enforcement or NCMEC
2. Select key frames from images to create a unique digital signature (hash)
3. Compare hash from key frames to a database of hashes from images classified as CSAM
4. Remove all images matching with hashes from database from internet. Manually review all remaining images
5. Add hashes of images identified as CSAM by manual review to database

AI image classifiers see and sort potential CSAM in **unknown images**

1. **Collect potential child abuse images** by crawling websites or scanning image before it's uploaded onto a user-generated site like YouTube
2. **Process and classify image** as CSAM or not, based on variety of factors including algorithms trained to detect nudity, age, body motion)
3. **Manually review** to confirm or deny CSAM classification; remove content from the internet give feedback to model so it can improve error rate

Therefore, AI moves in the direction of automating even more tasks – thereby freeing up human capacity to focus on high-priority issues.

AI tools are still being piloted to combat child sexual abuse online, however, they are showing significant potential in **preventing, detecting and prosecuting crimes** more efficiently and effectively – ultimately contributing to making the internet safer for children.[35]

Images are the core currency of online child sexual abuse but CSAM is surprisingly difficult to recognize. Indeed, the accuracy rate for CSAM tips reported by concerned individuals on hotlines is less than one third.[36]

Image analytics tools help **detect** abusive images and videos by analyzing underlying pixels or metadata. The technology works by creating a "hash" – a unique digital fingerprint – for each image. Hashed images that have been classified as CSAM by human analysts get compiled into a database. As tips of new potential CSAM are reported, new images are hashed and compared against the database of known images. If the image's hash exists in the database, relevant authorities can be contacted to take it down. Advanced forms of this technology can identify digitally manipulated images or images hidden within videos by making a "fuzzy match" to the original, hashed image. Microsoft's PhotoDNA tool has become the industry standard tool for hashing. AI moves from analyzing images to actually "seeing" and classifying them into categories based on patterns **detected** in the image.

Computer vision tools, which mimic the human eye, involve a number of overlapping technologies. An example is Griffeye's image classifier, which scans images for nudity, age and abuse. Through skin tone detection nudity can be determined – even in poor-quality videos. Separate programs extract facial features and perform spatial and textural analysis to determine if the face belongs to an adult or a child. Additionally, programs assess body motion to determine if videos are explicit. The classifier then generates an output score to determine if a file is CSAM or relevant to an investigation.

35 Examples include: Spotlight (by Thorn), which saves investigators more than 60% of critical search time daily; Freedom Signal (by Seattle Against Slavery), which reaches eight times more victims compared to in-person outreach; TraffickJam (by Marinus Analytics), which helped identify an estimated 3,000 victims in 2018; Project Arachnid (by the Canadian Centre for Child Protection), which to date identified more than 11.4 million images potentially showing abuse of children.

36 Based on interview with Canadian Centre for Child Protection regarding Cybertip.ca.

Language provides a key trail leading to all forms of abuse and is key to **prevention**. While encryption keeps an increasing amount of message content hidden, most abuse and advertising occur openly on the surface web. Much of the global volume of abuse can potentially be stopped at this initial language phase, regardless of how heavily coded or cryptic the language is.

Text and visual analytics tools search for keywords within noisy online web traffic and produce visualizations to help investigators and internet companies identify patterns of potential exploitation. These technologies are particularly useful in combating trafficking networks and identifying potential grooming behavior online. One leading tool, **Tellfinder,** combats sex trafficking by storing hundreds of thousands of online sex ads and visually grouping them by phone numbers, e-mails, and addresses so investigators can identify groups by the same trafficker. Data is populated as bubbles on a map so investigators can zoom in on a jurisdiction and scroll to see sex ads posted over time.

AI improves the ability to **detect** and respond to ever more faint signs of abuse online in a real-time and minimizes the need for human analysis. Natural language processing tools such as chatbots are used to read, understand and communicate using text, similar to humans. These capabilities are useful in decoding and intercepting abusive behavior that is being communicated online. An example is **childsafe.ai** which searches for sex trafficking on "hidden" popular commercial sex websites, or explicitly advertised on the dark web, using conversational chatbots to carry on text message conversations with buyers and providers of sexual services in order to gain additional information including pricing and location. Law enforcement is able to monitor and take action based on these automated conversations. Voice recognition tools are used to process and uniquely identify voices – a kind of "vocal fingerprinting". This technology is increasingly common in commercial products, including digital or home assistants, and can be used to identify perpetrators and victims in CSAM videos or audio communications.

FIGURE 2.3 **Evolving language tools**

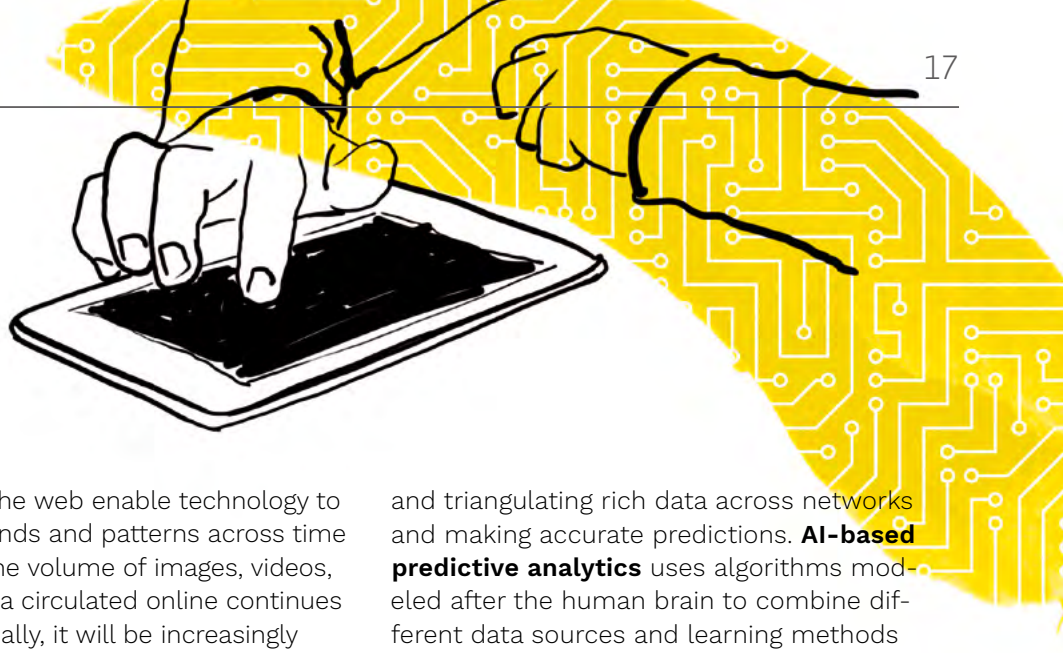
Splash pages deter users searching for CSAM using **keyword matching**[37]

1. Major search engines have lists of keywords to identify suspicious query patterns as they occur
2. User searches for CSAM using common list of terms known to perpetrator community
3. Splash page triggered – warning user about illegality of CSAM and offering anonymous options to seek help

Chatbots engage perpetrators online using **natural language processing** (NLP)

1. **Actively engage with perpetrator online** or passively respond to perpetrators and buyers of trafficking
2. **Carry on dynamic conversation** with NLP and content generation to mimic human text messages, including misspellings, humor
3. **Gather identifiable info on perpetrator** using game theory to carefully shift conversation towards topics with personally-identifiable info
4. **Take action online or offline using information,** such as triggering a splash page or coordinating sting operation with police at agreed upon fake location

37 A splash page is an introductory page that webmasters may use as a gate between the initial loading of the site and the actual site content.



Datapoints from the web enable technology to spot data flow trends and patterns across time and sources. As the volume of images, videos, voice and text data circulated online continues to grow exponentially, it will be increasingly important to efficiently and accurately analyze network trends to **detect** potential signs of online sexual abuse of children without necessarily having to examine the content of the data.

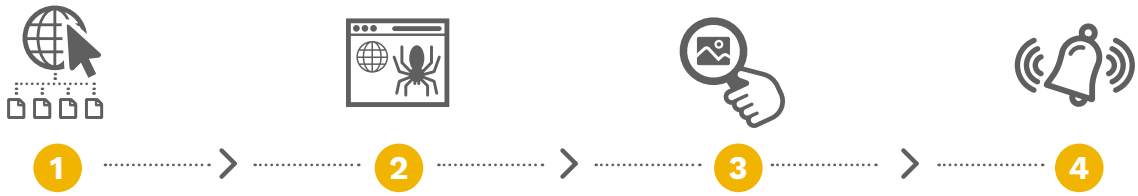
Network analytics tools comb through network data and use statistical techniques to identify useful trends and patterns that might point to criminal behavior across different platforms and help locate victims. These tools can also be used to analyze online interactions on social media accounts or web-traffic flows from dark to surface web. **Thorn's Spotlight** tool analyzes web traffic and other network data related to sex advertisements on escort sites to identify potential victims, monitor potential sex trafficking networks and generate leads for law enforcement.

Network analysis capabilities will become increasingly relevant as more and more content circulated online is encrypted due to very valid reasons of safeguarding individual privacy rights. AI is uniquely strong at analyzing and triangulating rich data across networks and making accurate predictions. **AI-based predictive analytics** uses algorithms modeled after the human brain to combine different data sources and learning methods to estimate the probability of a child being abused or the likelihood of someone being a perpetrator even when signals observable to humans are very faint. **Safeguarding Analytics,** for example, gathers data from both offline and online sources, such as the child's behavior on social networks, to make risk prediction about potential victims. Predictive, network-analytics-based AI models can also identify which perpetrators would be more likely to cause a network of abuse to collapse if targeted by an investigation – enabling scarce law enforcement resources to be more effectively prioritized. Predictive analytics powered by AI could help investigators quickly search and filter network activity to identify specific victims and perpetrators across the entire web as well as proactively generate leads to potential future perpetrators.

FIGURE 2.4 **Network analytics and predictive AI in action**

FIGURE 2.5 **Web-based tools in action**

Web crawling is used to proactively find CSAM across the surface and dark web



1 **Source URLs and images** via hotlines and follow links on confirmed CSAM sites

2 **Crawl surface and dark web,** scanning thousands of URLs and images per second

3 **Classify images** via hash, computer or human vision, confirming whether image is CSAM

4 **Notify hosts for take down.** No notices for dark web; identify links back to the surface web

Enabling tools work on the underlying infrastructure of the internet and devices to block access to confirmed CSAM or bad actors. For example, web-based tools are able to crawl across the dark web and escort websites, indexing all data they come across. Devices-based tools use software and hardware installed on children and employee's devices in order to filter or block access to specific categories of harmful content from being consumed or produced. Crisp for Kids and Teens is a filtering tool installed on children's devices that can dynamically filter inappropriate chat, text, image and video content in real time, such as grooming and cyberbullying attempts by bad actors. Crips's tool uses real-time behavior profiling to identify and remove bad actors from platforms including children's gaming websites and social media platforms.

FIGURE 2.6 **Overview of data analytics technologies**

Tool	Technology	Description	Common application
Image analytics	**Image hash values & databases**	▸ Creates a unique digital footprint for an image, called a hash, to be compared with databases of known CSAM hashes	▸ Identify "known" CSAM images across the web by sharing hash databases broadly across industry
	Metadata & context extraction	▸ Pulls out information and properties embedded in images beyond visual content, such as location, date, device used	▸ Identify victims or perpetrators based on contextual clues that recur across images
Text & visual analytics	**Text mining & analytics**	▸ Extracts information and insights from large quantities of unstructured text data by pattern recognition techniques, including flagging specific words	▸ Comb search histories or text scraped from websites for keywords associated with child sexual abuse
	Data visualization & mapping	▸ Combines data analysis and visualization, often using geographical data, to make large volumes of data digestible to humans and visualize patterns	▸ Track criminal activity, such as sex trafficking, across countries
Network analytics	**Network & predictive analytics**	▸ Uses statistical techniques to identify trends and patterns in network data, such as repeated criminal behavior across different platforms, locations and victims	▸ Analyze online interactions on social media accounts or web-traffic flows from dark to surface web

III. Mapping Current AI Solutions

AI's role in **preventing, detecting and prosecuting** online sexual abuse of children is emerging. The landscape of strategies for combating child sexual abuse online includes many solutions that rely on human judgement and intervention. At the frontline of combating the problem today are law enforcement agencies, government entities and non-profit cyber hotlines which largely follow manual, legally prescribed processes. For example, there are hundreds of global cybertip hotlines for reporting sexual abuse material, staffed by thousands of reviewers who verify reports of potential CSAM. One of the leading cybertip hotlines, Canada's Cybertip.ca, relies on three different human reviewers to verify a tip before making a CSAM classification.[38] The current approach ensures high accuracy but takes significant time and psychological toll on human reviewers. Even cutting-edge technology companies such as Facebook, Google and Twitter currently employ thousands of human moderators to manually review potentially harmful user comments and images on their platforms.

FIGURE 3.1 **Overview of AI technologies**

Tool	Technology	Description	Common application
 **Computer vision AI**	**Image classification**	▸ Automatically categorizes images as belonging to a group. The algorithms are trained by CSAM databases	▸ Conduct binary (CSAM vs. no CSAM) or continuous (age) classification. Classified images often verified by human reviewer at this stage
	Facial, object recognition	▸ Detects human features such as faces and body parts within an image and compares them to existing images to determine similarity	▸ Identify known victims and offenders in CSAM, as well as recurring inanimate objects in sex ads, such as hotel rooms
 **Natural language processing & voice AI**	**Natural language understanding**	▸ Uses algorithms to extract the words, syntax and semantics of a piece of text data	▸ Alert investigators to suspicious or abusive language online
	Natural language generation	▸ Artificially produces text that mimics human communication, can be trained to carry on conversations and mimick tone of individuals	▸ Use chatbots to engage perpetrators on online forums and messaging apps
	Sentiment analysis	▸ Extracts subjective opinion or sentiment from text, video or audio data	▸ Detect subtle signs of distress or abuse in potential victims in an online environment
	Speech recognition & voice analytics	▸ Processes and uniquely identifies the "vocal fingerprint" of individuals	▸ Identify perpetrators and victims in CSAM videos from their voice
 **Predictive AI**	**Data mining for patterns & trends**	▸ Explores large datasets for insights, patterns or relationships between variables	▸ Develop models to aid in complex sex trafficking investigations that involve various types of unstructured data
	Early risk identification	▸ Assigns a risk score to a piece of content using algorithms trained on past data and continuously fed with new data to improve accuracy over time	▸ Use as predictive policing tools. Can be used to identify potential victims or perpetrators

38 NetClean. "How Web Crawlers Can Help Find Child Sexual Abuse Material". Technical Model National Response. 2018.

In addition to manual approaches, there are hundreds of technology tools being used today to fight online sexual abuse of children. The most widely used tools have been developed by bringing together technology expertise from the private sector or academia with specialized knowledge from NGOs and law enforcement agencies working at the frontlines. They are primarily analytics-based and reactive, requiring human intelligence and decision-making to take action.

A small, promising subset of tools leverage AI. A mapping of around 50 emerging solutions captures the current maturity of AI in the global fight against online sexual abuse of children. These have been categorized and described in accordance with their role in **prevention, detection or prosecution**.

The mapping reveals the geographic focus of AI solutions is still limited, though the nature of AI technology bodes well for global expansion, provided supportive legal and policy conditions exist. **Canada, UK and US** lead the way in using AI to fight this crime with strong government initiatives and numerous high-profile, cross-sector partnerships bringing together leading technology players and NGOs.

The most visible players using AI in the global fight are **nationally-based child sexual abuse hotlines** which typically start by focusing on a safer, cleaner internet in their own country and then proceed to push for global action and collaboration on the issue. The borderless nature of internet-enabled child sexual abuse means that threats to children arise from all geographies, making national solutions insufficient. The US-based National Center for Missing & Exploited Children (NCMEC) runs a CyberTipline that functions as a global clearinghouse for CSAM, making its reports available to US law enforcement and more than 100 law enforcement agencies worldwide. Sharing databases of CSAM images and hashes is central to the challenge of identifying victims and perpetrators, given that 85% of CSAM features unidentified victims.[39]

Beyond engaging with law enforcement, these hotlines engage internet platform companies to voluntarily scan their platforms to keep them clean of abuse. For example, Canada-based Project Arachnid, which started by crawling through sites reported to Cybertip.ca, now reaches out to the internet industry with an API that will crawl on their systems. The UK-based Internet Watch Foundation plays a similar leading role in Europe of coordinating national and international responses to CSAM reports by engaging directly with internet industry players.

Beyond these national initiatives, there are several **bright spots for international collaboration enabling the use of AI**. INTERPOL and Europol are active drivers of collaboration between law enforcement agencies across borders. INTERPOL maintains a global database of CSAM gathered from hotlines and law enforcement agencies that it continually updates to aid in global victim identification. INTERPOL is currently pushing for increased data sharing to expand global CSAM databases as well as more consistent practices for annotating images, which is essential for ensuring AI solutions can be developed and continuously improved in supervised learning environments with quality training data. In 2015, the United Nations Interregional Crime and Justice Research Institute (UNICRI) launched its program on AI and robotics with a focus on exploring how they can contribute to a future free of violence and crime. UNICRI supports international, cross-sector research and the development of partnerships to fight cyber and traditional crime with the help of AI. As a partner of INTERPOL in the organization of the annual Global Forum on AI for Law Enforcement and a member of the Global Partnership to End Violence against Children, UNICRI could be a promising multilateral partner to develop global AI-powered solutions to help law enforcement agencies **detect and prosecute** online-enabled crimes involving sexual abuse against children.

The mapping of AI solutions shows **collaboration across sectors is a success factor** – particularly when private sector technology companies partner with NGOs and law enforcement agencies on the frontlines to co-create tools that target the most visible forms of online abuse.

39 Based on interviews with Canadian Centre for Child Protection regarding Cybertip.ca

FIGURE 3.2 **AI solutions across prevention, detection and prosecution**

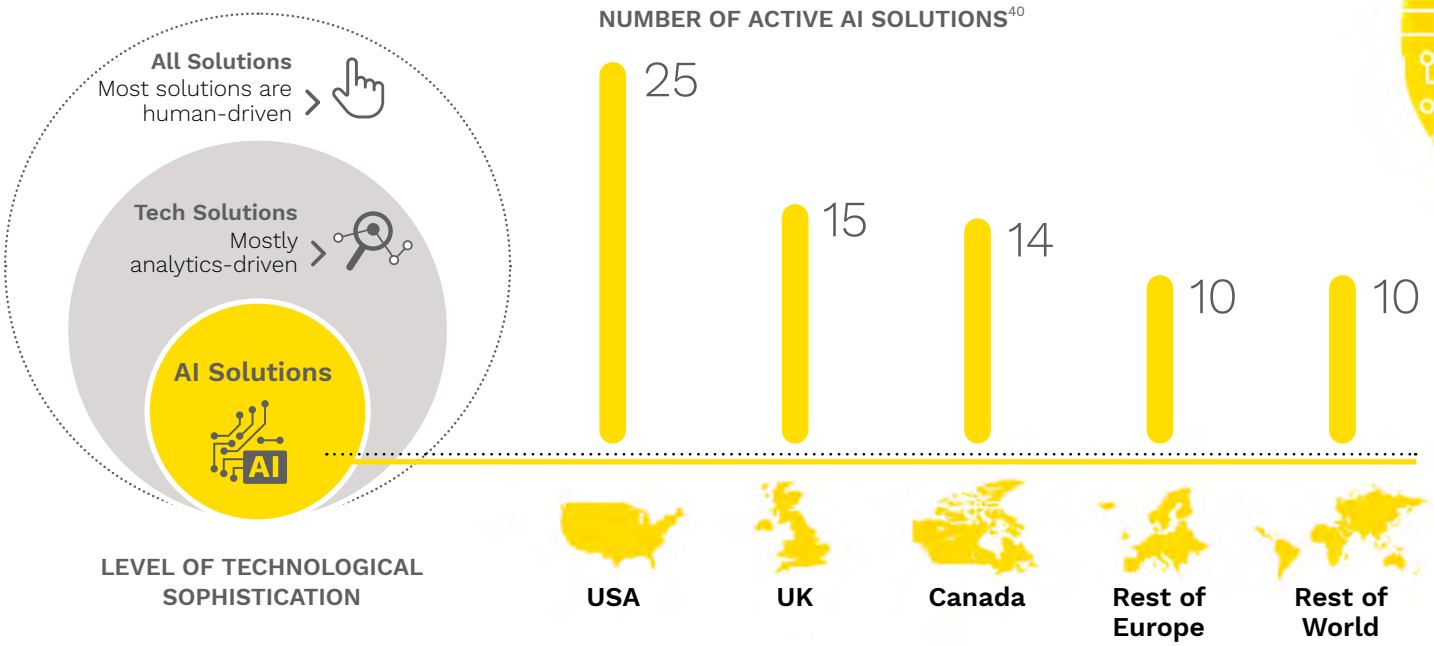
Prevent

Detect

Prosecute

What	How	Solutions & organizations	
Protection of children from vulnerable behavior online	▸ Identify potential for cyberbullying and child abuse in conversations, real-time harmful image or video sharing, and intervene in an automated way ▸ Install parental software or device-level hardware to filter abusive content or prevent consumption and production of harmful content	Deterrence	Thorn
		Ovoty pp	Privately
		SafeToNet	SafeToNet
		childsafe.ai	childsafe.ai
		Intercept Chatbot	Freedom Signal
		Social Network Analysis	Safeguarding Analytics
Proactive interception of online grooming attempts from sophisticated predators	▸ Identify predatory user behavior on social media and child-friendly sites using network analysis, followed by automated flagging and banning	Rapid Safety Feedback	Eckerd Connects
		Predpol	Predpol
		Crisp for Kids and Teens	Crisp
Reactive deterrence of perpetrators in search processes	▸ Install automated messages and search engine filters for curiosity-driven, surface-web CSAM browsing on major search engines, ISPs, commercial escort sites		
Image-based tools to flag CSAM for human review	▸ Pre-classify CSAM for human review, using nudity and age detection technology, followed by takedown process ▸ Proactively source potential CSAM via web-crawling and scraping for content "at its source" before it is widely circulated	ProActive	NetClean
		Human reviewers	Internet Watch Foundation
		Project Arachnid	Canadian Centre for Child Protection
Non-image, text or data-based tools to detect signs of abuse	▸ Detect signals via text-based analysis, contextual analysis of online patterns of posting and behavior indicative of trafficking ▸ Use data hubs to detect signals via analysis of financial or communications flows between parties online	Content Safety API	Google
		Qumodo Classify	Quomodo
		Brain CSA classifier	Griffeye
		CEASE.ai	Two Hat, ImageVision
Advanced image and non-image-based tools to tie images to specific individuals or locations (victim and perpetrator)	▸ Identify victims via advanced facial recognition including "fuzzy" matches of non-identical images ▸ Identify perpetrators via image-based analysis (e.g. scene details like hotel room) or recurring communication patterns	STOP App	Stop the Traffik
		TraffickCam	Exchange Initiative
		Minerva	Global Emancipation Network
Tools to translate digital signal into physical location of victims and traffickers	▸ Use decoys and automated chatbots to capture location-specific information from victims and traffickers ▸ Analyze trends using signal from images, metadata and text communications to identify trafficking networks, patterns and hot pots traced to IP addresses	Spotlight	Thorn
		Tellfinder, Datawake, Dig	Darpa Memex
		Trafficam	Marinus Analytics
		Traffik Analysis Hub	Traffik Analysis Hub
Enhanced investigative tools with the ability to sense, search and build digital cases quickly	▸ Improve automated search and filtering capabilities ▸ Implement digital forensic tools to track and gather evidence of online perpetrator behavior ▸ Use tools to visualize and simplify trends and patterns, drawing on live and historical data from advanced data hubs	Intercept	Seatlle Against Salvery
		Minerva	Global Emancipation Network
		Toolkit	Thompson Reuters, Mekong Club

FIGURE 3.3 **Landscape of current AI solutions**



Most AI tools have focused on CSAM detection and removal, together with tools that aim to gather and analyze global data on online child sexual trafficking cases. This concentration is driven by the fact that these traditional forms of online child sexual abuse are more publicized, and there is a greater awareness of the role that technology, social media and internet providers can play in facilitating these forms of abuse. This has led to heightened calls for these players to take action to combat the problem. Google, Facebook and Microsoft have launched CSAM-focused, AI-powered tools for their platforms that are also made available to all internet players as an API. While there is strong voluntary engagement from the largest tech players, in part due to mounting public scrutiny, more commitment is required from the broader circle of interactive platforms (such as the multitude of gaming and social interaction platforms that attract children) to ensure they take action to keeping sites "clean" by proactively **detecting** and eliminating child abuse.

There are numerous efforts to harmonize takedown practices of child sexual exploitation and CSAM from online platforms, including a flexible Industry Hash Sharing Platform as well as standardized protocols for processing CSAM images. An emerging standard called VICS (Video Image Classification Standard) was created by a coalition of CSAM investigators, victim identification specialists, application developers and scientists to make it easier to exchange comprehensive sets of hashes domestically and internationally without having to touch or manipulate the data – a big step in using collaborative data to identify known images – including with AI.

Efforts to collaborate on fighting the new, fast-growing forms of online child sexual **abuse using AI are just beginning.** To tackle online grooming, partnerships are emerging to better safeguard children on popular youth platforms. For example, Operation Game Over in 2012 saw Microsoft, Apple, Electronic Arts, Disney Interactive Media Group, Warner Brothers and Sony remove more than 3,500 accounts of registered sex offenders from online video game platforms such as Xbox Live and PlayStation thanks to AI tools.[41] An increasing number of cross-sector "hack-a-thons" and collaborative "virtual labs" are being organized to find proactive approaches (many of which are AI-enabled) to identify and combat online grooming. Additionally, for-profit AI-powered solutions that help concerned corporations, platforms, and parents keep their devices "clean" of CSAM are emerging.

40 One AI solution can be active in more than one region.

41 Joseph Goldstein. "Video-Game Companies Agree to Close Sex Offenders' Online Accounts". New York Times. April 2015

IV. The Path Forward for AI

There is significant untapped potential for AI across the current landscape of human and analytics-driven approaches to fight online sexual abuse of children. By building on a strong base of proven capabilities, AI can be leveraged in three ways to play a pivotal role moving forward.

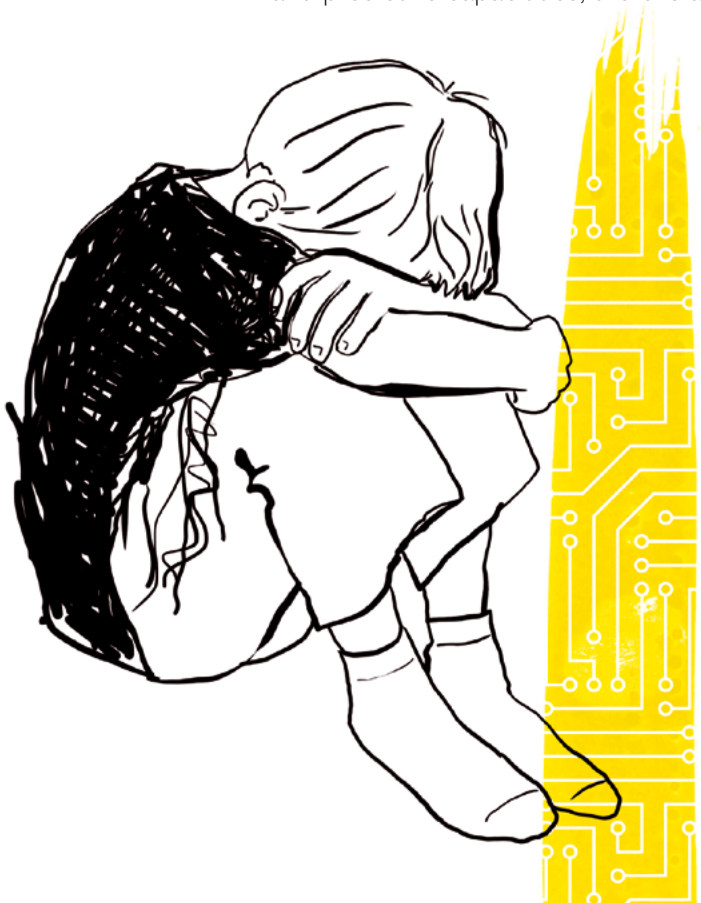
1 **Upgrade** existing manual and analytics solutions with the latest AI capabilities

2 **Shift** development of new AI tools from detection to prevention and prosecution

3 **Expand** reach of AI solutions to new geographies and to new forms of abuse

1. Upgrade existing manual and analytics solutions with the latest AI capabilities

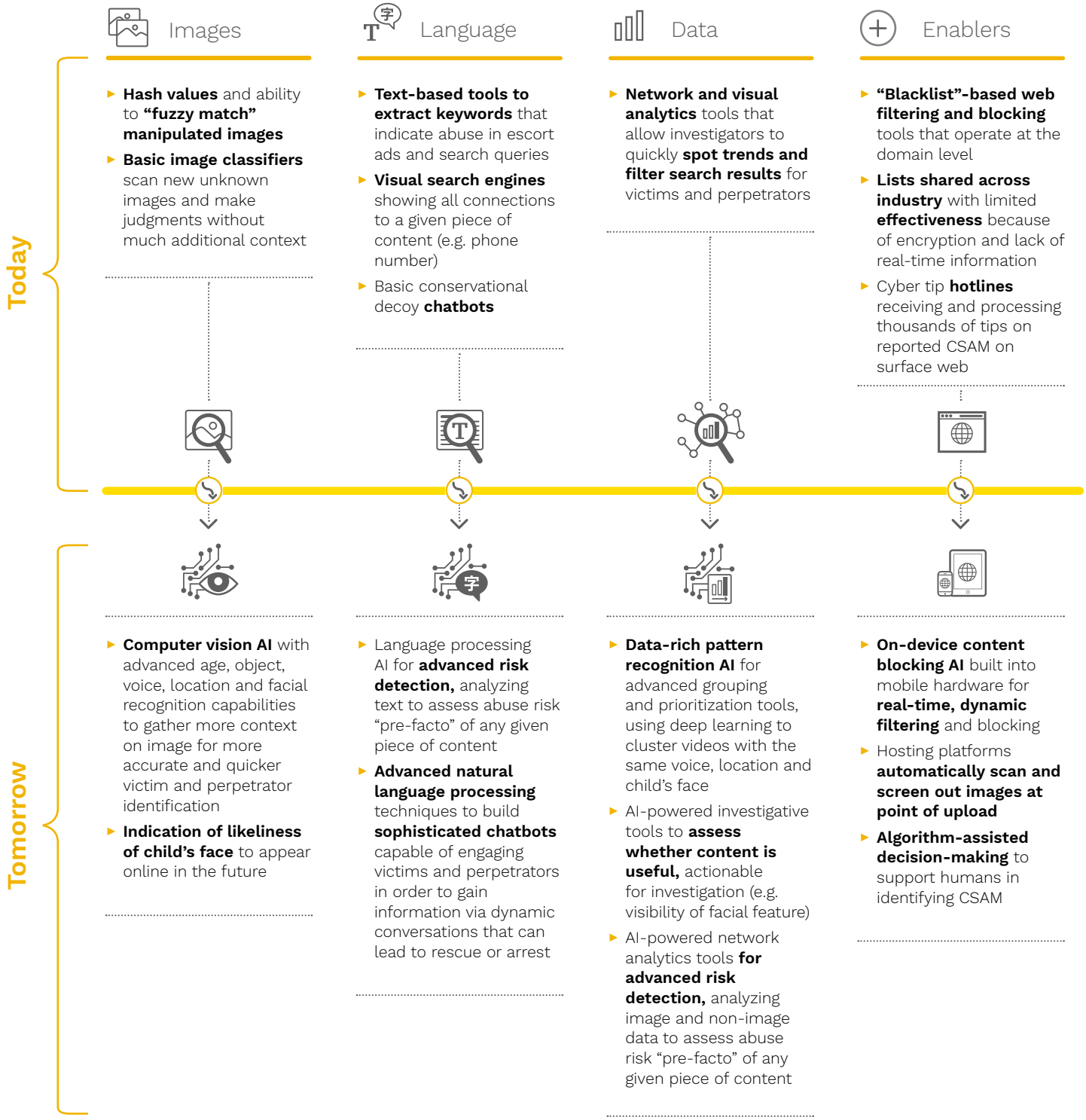
AI technology can transform many of the labor-intensive, reactive approaches to combating the problem. Across image, language and predictive capabilities, there is a clear path for AI-powered tools to automate the most time-consuming and complex tasks that human reviewers and investigators complete today. Using AI that is faster, cheaper and more accurate than current tools will allow the industry to more adequately confront the growing magnitude of the problem.



Upgrading tools with AI capabilities does not take humans out of the decision-making loop. Rather, algorithms will allow humans to process many more cases as they prioritize material and remove many tedious and time-consuming tasks. That said, AI enhancements should ensure they do not simply improve the scale of **prevention, detection and investigation** efforts at the expense of quality or accuracy. Indeed, the proliferation of AI-supported solutions should not coincide with an exponential growth of innocently charged people or of children being prevented from using interactive platforms on the internet. All AI solutions need to be carefully developed and tested to avoid collateral damage in secure environments before being piloted, refined and scaled up.

FIGURE 4.1 **Disrupting abuse with Next Generation AI solutions**

The current forms of online sexual abuse of children today which include but are not limited to the spread of CSAM, online grooming, cybersex trafficking and live-streamed videos all stand to be disrupted by innovations in the way we integrate tomorrow's AI technology into the solution.



Next Generation AI solutions will reshape the way online sexual abuse of children is **prevented, detected and prosecuted**:

- From reactive solutions that remove evidence of past abuse to **proactive solutions** that focus on preventing online abuse and **prosecuting** "top of the pyramid" predators responsible for driving most of the abuse online
- From surface web-focused searches to **entire web searches, including all relevant platforms and the dark web**
- From hotlines and criminal investigations driven by humans, to **algorithm-assisted decision-making, tools and case-building** which have the capability to search and act on abusive material in an increasingly automated way (with human oversight)
- From basic analytics tools such as image hash lists being the global industry standard to a powerful, **universal AI image classifier** used across geographies to determine CSAM, while referencing a **shared global database of all known CSAM victims and perpetrator imagery and other contextual information** to aid in identification
- From text-based tools that search for keywords to **advanced network analytics tools that can detect and flag suspicious user behavior** and share information across platforms
- From web-based tools that crawl the internet searching for CSAM to **built-in, on-device AI powered operating systems that prohibit CSAM from being self-produced or consumed**
- From **post-factum** criminal investigations **after the abuse has already** happened to **real-time takedowns** of the channels used to enable that abuse
- From curbing the rise of CSAM and online grooming, to **full eradication** of such criminal activities in any environment accessible to children

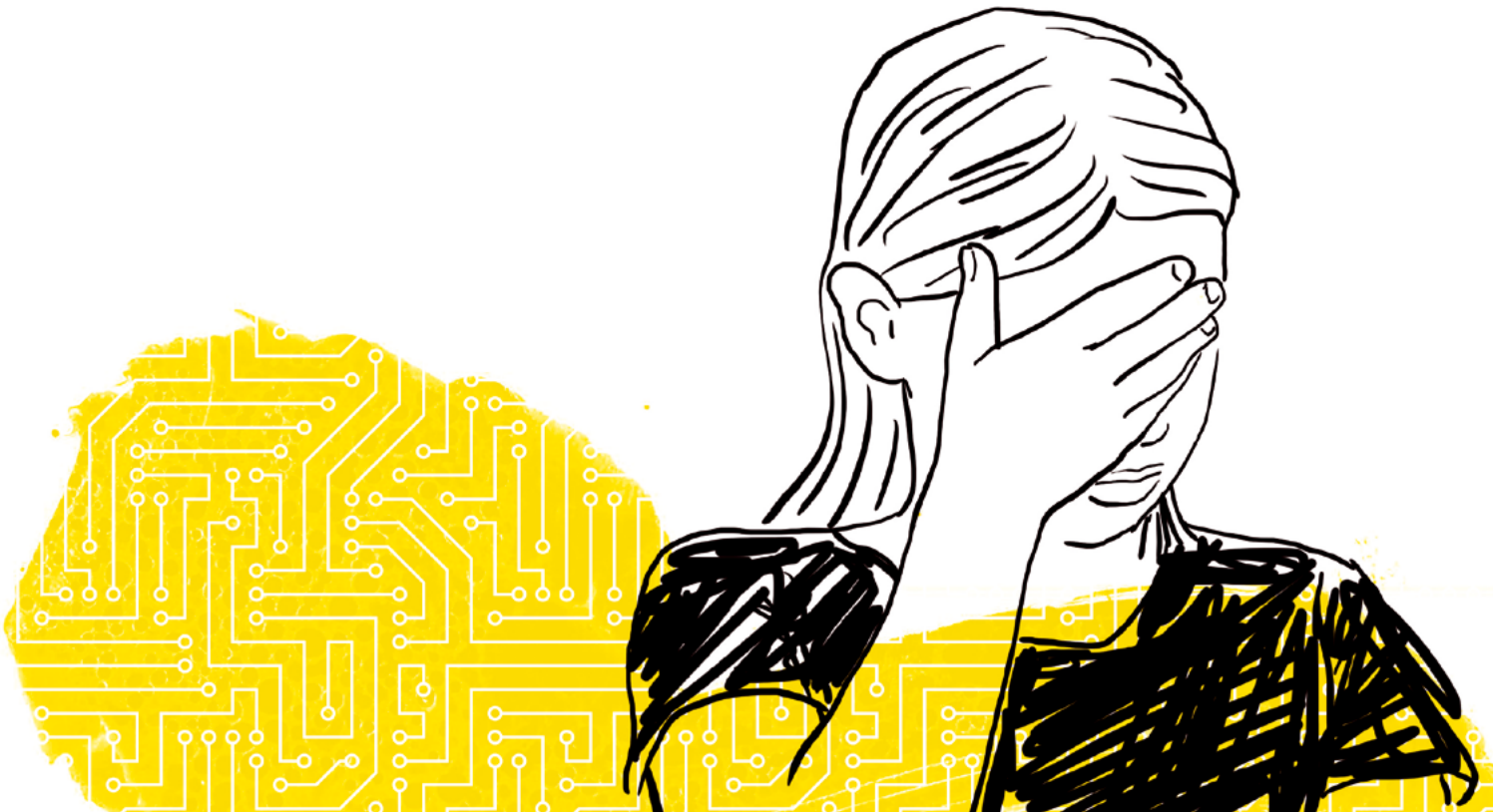
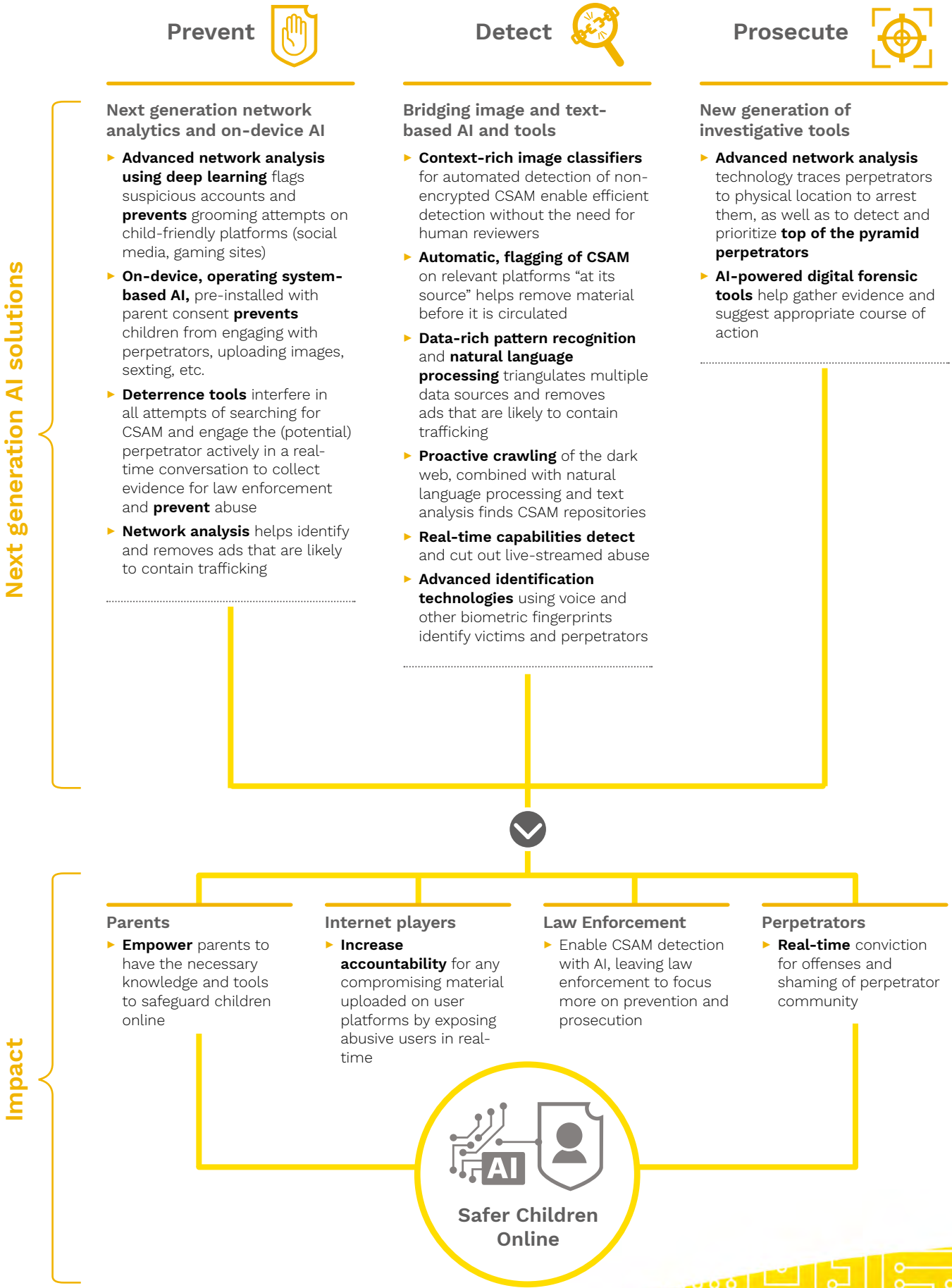


FIGURE 4.2 Mapping Next Generation AI solutions by use case



2. Shift focus from AI tools on detection to prevention and prosecution

A. The future of prevention requires next generation network analytics and on-device AI.

Prevention is challenging because parents are often not aware of their children's behavior online and of the potential risks. Current blocking and filtering programs are ineffective to real-time threats and encrypted environments. Further, these tools do not keep children from engaging in risky online behavior.

To address the challenges of encrypted communication, **sophisticated network analysis, powered by deep learning techniques,** is required to spot signs of abuse in user interaction patterns and automatically block online grooming attempts. To develop AI solutions, interactive online platforms frequented by children, such as social media and gaming sites, should ideally be able to legally capture and share data with law enforcement entities in such a way that safeguards privacy rights of platform users for the sole purpose of developing accurate, AI-powered network analysis solutions that can better **detect** signs of grooming and other forms of online abuse to better safeguard children.

A more comprehensive solution, and one that addresses the ineffectiveness of current blocking and filters, is to invest in **on-device (operating system) powered AI** solutions that can identify potentially compromising videos, photos, voice and text messages captured at the source. Such solutions would run on the device itself – rather than the cloud – and could engage children with friendly, AI-powered chatbot messages to warn them of the risks of uploading compromising data online or having a blocking function that **prevents** compromising material from being captured and stored on the device in the first place. Sexting, for example, could be **detected** and **prevented** as soon as a photo is taken. On-device AI solutions that incorporate the latest image and language AI capabilities to identify, block or remove potentially risky content would require collaboration by Apple, Google and Microsoft – whose operating systems jointly run 97% of global internet-connected mobile devices.[42] Additionally, any such solutions would need to be developed in accordance with national laws governing children's rights to self-expression and protection. The major advantage of investing in such on-device solutions for children is that they help fight the problem at its source by reducing the number of self-generated images landing online in the first place.

B. The future of detection requires bridging image and text-based AI and disrupting abuse at its source before it can circulate.

The magnitude of CSAM, including the proliferation of new forms of CSAM such as "deep fakes" – in which, for example, a children's face might get superimposed on another person's nude body – means that human inspection of images will be an increasingly less viable strategy for combating the problem. Future solutions should focus on leveraging AI to **detect** CSAM across the web, and more importantly, on platforms where it is uploaded and spread in the first place. AI can help with both the reactive and proactive **detection** and removal of online CSAM.

Advanced computer vision technology, when trained with quality data from global CSAM databases, can produce image classifiers capable of identifying new, previously unknown CSAM at levels of accuracy comparable to or greater than human inspection. For more difficult, borderline cases that require more data points from different sources to make the determination, an exciting area for future investment is to **bring together the strengths of image AI and language-based AI** to help solve the problem of weighing context in the classification of CSAM.

42 Rustam Aliyev. "How Artificial Intelligence Can Help Protect Children". Purify Foundation. November 2018.

To deal with **detection** of abuse occurring across the surface web, the challenge is finding the digital signal within the noisy online marketplaces, forums and websites with thousands of simultaneous post, transactions and chats where sex trafficking can hide in plain sight. Going forward, the **advanced pattern recognition capabilities of predictive AI** need to be leveraged to **label and group (potentially) abusive content** based on recurring location, voice, image and text patterns across platforms and over time.

Voice AI technology has accelerated with commercial applications and can be used to uniquely identify individuals in any video or audio recording using their voice "fingerprint". Similarly, **"deep fingerprint"** technology is being developed to embed unalterable fingerprints into images and videos that are uploaded to the internet. This technology could allow online platforms to **authenticate whether an image or video has been altered from its original format**. The frontier of identification technologies is using **biometric AI** to identify victims of human trafficking. It is now possible to scan a child's face and **create a unique identification based on the child's iris** in a few seconds. If a child is abducted, iris scanners on surveillance cameras in public places such as airports could match irises to original images.

Another area that AI can help with is to **disrupt the abuse or crime in real time –** such as by halting a session of live-streamed abuse or intercepting an intended financial transaction for child sex trafficking. There is promising AI technology from the commercial world that is able to **detect suspicious financial transactions as well as illicit video streams** within the vast pipeline of online data. If trained to look for signs and patterns of digital abuse, such as pairing a digital payment for a certain fixed amount followed by Skype sessions, AI models could **detect** and disrupt abuse in real time.

C. The future of prosecution requires a new generation of AI-powered investigative tools that can predict and prioritize.

Prosecution is incredibly challenging in the digital era due to the massive scale and scope of modern cybersex crimes. Law enforcement officials are largely relying on a past generation of manual and analytics-based tools to build cases against increasingly sophisticated and prolific perpetrators who are getting better at covering their digital tracks and typically operate in anonymous online environments. There is a pressing need to upgrade investigative tools with AI capabilities so investigators can better search, filter and build cases in the digital era while also reducing the psychological toll on human investigators by appropriately labeling and prioritizing the material they need to sift through.

A vision for the future would be to have **AI-powered digital forensics tools** that can help law enforcement sift through vast amounts of online activity data to identify potential patterns of abuse and trace them back to individual perpetrators. The biggest gap currently is tools that can dynamically track the movement of sex trafficking networks across different geographies. Unlike advances in sharing databases of hash values, the data required to catch sex traffickers is temporary, context-dependent and does not move easily across borders and sectors. Fortunately, **advances in network analysis** to **detect** trafficker movements can be **paired with advances in image and context recognition** to help tie online signs of abuse to physical locations and specific individuals. AI is capable of recognizing recurring trends across images and posts, such as the background of the scene in different videos, or trends in the types of language and images used by traffickers to advertise services online. These AI capabilities have significant potential to identify and track down perpetrators, especially across regions with multiple law enforcement agencies – each with their own databases – such as in Europe.

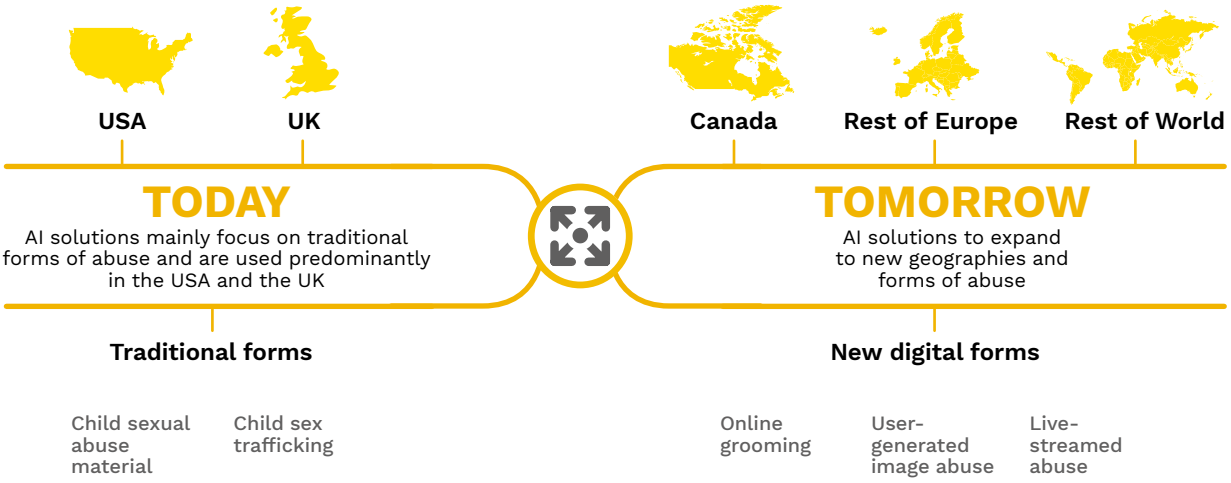
The future of AI-powered investigative tools is to help investigators quickly assess whether a given piece of content is relevant and actionable for prosecution. Advances in risk detection powered by deep learning techniques increasingly make this possible. For any given image, text fragment or phone number, AI will soon be able to predict the likelihood of abuse tied to the respective piece of content. Additionally, it can indicate to investigators whether the content is useful in an investigation – for example, by quickly calculating what percentage of facial features are visible. Rather than sorting through thousands of images and posts and subjecting themselves to a potentially huge psychological burden while doing so, investigators can use these tools to prioritize reviewing the most relevant material as they build cases against child sex traffickers.

Another area for potential high impact in investigations is to **develop solutions that help identify and arrest the "worst of the worst" of online perpetrators** – those who produce and disseminate CSAM. This remains a challenging gap in combating online abuse, but advances in predictive risk AI can be leveraged to identify and build cases against those who drive the CSAM ecosystem of online repositories and communities.

3. Expand the reach of AI solutions to new geographies and forms of abuse

A key priority in the future should be to **expand North American AI solutions to Europe and other geographies** where children are at high risk of being vulnerable to grooming and online sexual exploitation. There is significant potential for technology to be deployed in the fight against child sex trafficking in Europe which hosts the majority of CSAM globally – but this will require working through challenges related to the uncertain application of Europe's General Data Protection Regulation (GDPR) to sharing information on online child sexual abuse and the fragmented setup of multiple national law enforcement agencies. In lower income countries, where live-streamed abuse is largely concentrated, there is an urgent need to deploy AI-powered tools to curb the growth of new forms of online abuse.

FIGURE 4.3 AI solutions should expand to new geographies and forms of abuse



To move AI forward, several key barriers need to be overcome.[43]

Legal and policy barriers to accessing and sharing annotated datasets of CSAM and other data related to online sexual abuse of children online is currently impeding the development of global AI solutions, which requires quality, representative data to train AI models. Today, law enforcement and NGOs are hesitant to share data with actors willing to develop AI-based solutions – in large part due to the transnational, global nature of the crime that is in conflict with national laws prohibiting sharing sensitive data across national lines. This uncertain legal environment creates significant barriers for investment in solutions that deploy advanced AI technology at a more global scale. **Data on hashes and images should be able to be used and shared across jurisdictions** facing the problem as "global asset". More broadly, legislation has failed to keep pace with the digital age, and there is a patchwork of national, regional and global standards and laws that needs to be revised and harmonized. The Online Harms White Paper published by the UK Government is a proposal in the direction of responsible, forward-thinking action on regulating risks in the new online environment.

The lack of a unified, clear legal definition of CSAM and an approach for combating it across industry or law enforcement globally presents a key barrier for technology players and other actors to invest in AI-powered solutions. Thirty-five countries have no laws criminalizing CSAM, and of those that do 76% lack a definition, making **prosecutions** very difficult and creating a zone of impunity for perpetrators. Work needs to be done to establish consistent definitions, **a single standard framework for the classification of CSAM, and standardized toolkits** of technologies that can be used.

The rise of new technologies, from encryption to virtual and augmented reality to the decentralization of the web (dark web, cryptocurrency), promises to intensify the problem and make it virtually untraceable. The global push for end-to-end encryption will make much of the online activity of perpetrators invisible to investigators.

Law enforcement and **AI technology need to adapt to a new, all-encrypted world**. Fortunately, anonymity does not render AI useless in the fight against cybercrime. With device-level AI to protect children and more sophisticated network analysis tools to identify suspicious data flows there are many ways to circumvent the challenges of encryption and decentralization.

Securing cooperation across the landscape of internet platforms is another key barrier, both in responding to takedown notices as well as proactively scanning and eliminating abusive material from platforms as soon as its uploaded – similar to Facebook and Google today. For example, the Netherlands hosts nearly half of the world's CSAM in part due to the reluctance of hosting platforms to take down material unless compelled legally. The largest internet players have gradually made changes, but significant effort is required to **gain the cooperation of the next layer of platforms and websites that currently prop up child abuse**. The shuttering of the popular escort site Backpage.com in 2018, which facilitated sex trafficking, was a success that needs to be built upon globally.

Finally, there is a general **lack of awareness about the problem and how AI technology can help**. There are many misconceptions regarding the extent and nature of child sexual abuse online – especially for emerging forms of abuse that are poorly understood or measured. There is even more confusion about the nature of AI technology and the role it can play in protecting children from sexual abuse enabled through the internet.

43 This section draws on recommendations from: Child Dignity in the Digital World. "Child Dignity Alliance: Technical Working Group Report". 2018.

V. Call to Action

Bringing the full strength of AI technology to the fight against online sexual abuse of children cannot be achieved without the full participation and concerted efforts from key stakeholders. It needs six essential elements to create a world in which the **internet and online connectivity are genuinely a force for good for children worldwide** – thanks to the safeguarding power of AI.

1. Share knowledge and increase collaboration among stakeholders

Technology players need to bring in their AI know-how; law enforcement agencies their intelligence on perpetrators and criminal activity trends; and civil society actors their deep understanding of issues faced by victims to ensure solutions respect children's rights.

2. Establish new forms of collaboration across sectors and borders

The different stakeholder groups are required to collaborate across borders and across sectors with an unprecedented determination and scale to beat the massive global communities of perpetrators online.

3. Redefine legal frameworks and cooperation agreements enabling secure use and sharing of data

New and harmonized legal frameworks are key to facilitate cooperation between multiple stakeholders and to enable secure data transfer respecting data privacy regulations, such as GDPR.

4. Allocate more resources to develop and expand AI solutions

Public authorities and private actors need to closely collaborate and jointly invest not only in tools that improve detection, but also in tools that go beyond and disrupt abuse online the moment it occurs.

5. Increase awareness and understanding of the severity of the problem and its many forms

More awareness among the general public around the magnitude of online sexual abuse of children and available solutions is required to increase the understanding of its new forms and facets.

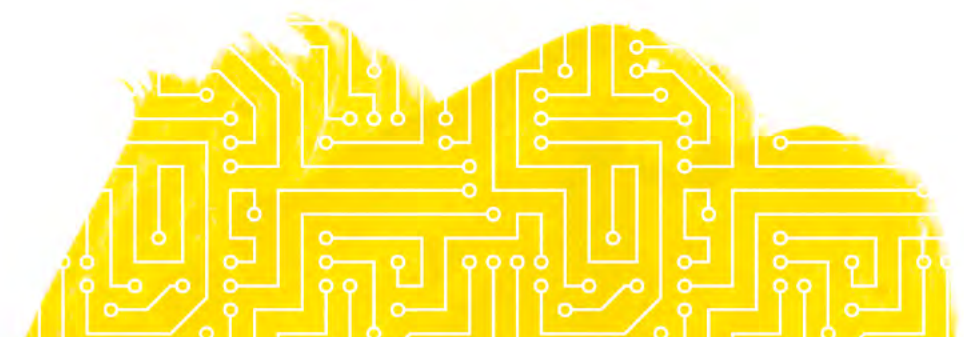
6. Invest in the development of enhanced digital skills for both law enforcement and civil society

Only with sufficient expertise, such as data science skills, will law enforcement be able to absorb the opportunities technology brings to the table. Further, non-profit actors and parents need digital skills to make informed decisions.

FIGURE 5.1 Key priorities and stakeholders

Action required	Key stakeholders	Specific steps, roles, responsibilities
Players with technology capabilities should proactively bring the latest AI technology to combat online child sexual abuse	**Technology and AI-focused industry**	▸ Apply technologies that are proven commercially to the key gaps highlighted in this report by **co-creating technology solutions with NGOs and law enforcement** who are closest to the problem (for example, Google's Content Safety API, Microsoft's PhotoDNA) ▸ Move from a reactive approach of developing tools for law enforcement to a **proactive approach of collaborating with public authorities in developing and implementing solutions** for major online platforms to ensure CSAM and exploitation can be quickly detected, reported and removed ▸ Shift **focus to new forms of online child sexual abuse** before they grow out of control, as well as tools that move beyond **detection** and focus on disrupting instances or patterns of abuse online ▸ Work with **hardware and device manufacturers** to solve blocking and filtering shortcomings with on-device AI
	Internet industry (interactive platforms, internet service providers, websites)	▸ Make better **use of data collected** to proactively identify high-risk accounts (victims and perpetrators) and analyze behavioral patterns ▸ **Work with law enforcement** to develop solutions to legally disrupt abuse in an efficient (and if possible automated) way ▸ Respecting national legal parameters, **share operational and intelligence data** about confirmed criminals with other platforms, and create a "blacklist for bad actors" in addition to current URLs and images blacklists
Players at the frontlines need to push for more technology partnerships and a better legal framework to more efficiently and effectively address the problem	**Law enforcement agencies** (at all levels)	▸ Make the **case to legislators that AI can be highly effective** in supporting the work of law enforcement and that more legal provisions should be made to allow the use of AI in **prevention, investigaton and prosecution** efforts – particularly related to cybercrimes against children. To do this, work with trusted AI industry players on developing pilot solutions that show the power and accuracy that AI can have in classifying images, voice, text and network data to identify potential patterns of criminal activity ▸ Where possible, **invest in AI tools and other technologies** that not only make investigation and **prosecution** efforts more efficient but can also significantly reduce the psychological burden of human analysts and make their workplaces safer ▸ Work with AI industry players on developing **solutions for the greater challenge of finding ways to uncover abuse in private homes**; individual "lone wolves" are more difficult to **prosecute** compared to a crime ring ▸ Actively collaborate with other law enforcement agencies across the globe; look for opportunities to **harmonize approaches and technology tools used** ▸ Help **educate public on gravity and scale of crime**, especially new forms of live-streamed and image abuse
	Country governments & legislators	▸ Create the necessary **legal accommodations to facilitate effective cooperation between knowledgeable industry players and public authorities** to develop and pilot solutions using AI to better **prevent, detect and prosecute** crimes against children ▸ Respecting GDPR provisions, develop **model data transfer agreements** to facilitate the sharing of CSAM, trafficking and other relevant data between law enforcement and vetted industry players developing AI solutions ▸ Establish **guidelines for the design of devices, platforms and online environments** that ensure safety of children, especially in regard to new emerging technologies (for example, the UK's Online Harms White Paper) ▸ Empower law enforcement with **funding and tools appropriate** for the current and growing scale of the problem; ensuring sufficient data science expertise is available to law enforcement to operate and refine tools ▸ Make **legal provisions for therapy** to be made available to individuals who have been identified as potentially abusive to children online; ensure sufficient funding is provided for this important preventative measure

Action required	Key stakeholders	Specific steps, roles, responsibilities
Players with financial resources should catalyze piloting and developing AI solutions for new problem areas and geographies	**Socially-minded investors**	▸ Provide **early-stage funding** for pilot projects between AI technology companies and relevant law enforcement, NGO or government partners ▸ Focus on **"whitespace" geographies** (Europe, Africa, Southeast Asia) and areas of the **problem** (prevention)
	Socially-minded corporations	▸ For device makers and interactive online platforms, position **"child safety" as commercial advantage** ▸ Ensure **participation of payment services providers** (for example Western Union, PayPal) to help identify financial transactions related to sex trafficking and live-streamed abuse ▸ Offer low or **no cost safeguarding tools** to parents and CSAM scanning and **detection** tools to websites and platforms to keep networks clean of abuse material
	Academic and research institutions	▸ Take advantage of legal provisions that enable **scientific collaboration with law enforcement to spearhead research projects** on the development of AI solutions against child sexual abuse alone ▸ Bring **latest research on AI technologies to pressing, neglected areas of the problem** of online child sexual abuse in collaboration with public authorities and/or private sector technology players
Civil society must continue to build awareness and demand that public and private sector actors take action	**Domestic and global NGOs**	▸ Build **coalitions across public sector, technology firms, academia and law enforcement** (for example, WePROTECT Global Alliance, Child Dignity Alliance, Traffik Analysis Hub) ▸ Launch **public campaigns to educate** children on the safe use of the internet and raise awareness of risks of online abuse with parents and the preventative measures that can be taken ▸ Launch **awareness campaigns around the magnitude of the problem** and technology as the key tool for the fight ▸ **Standardize approach to handling and annotating CSAM** in line with global best practices to facilitate collaboration with domestic partners and other NGOs globally
	Multilaterals and global alliances	▸ Establish **cooperation frameworks with universal standards for legal sharing of data and intelligence** between law enforcement agencies globally, as well as between law enforcement agencies and trusted private entities with promising technology to bring to the fight ▸ Make **CSAM held by law enforcement agencies a global data asset** that is not restricted to single jurisdiction or agency so AI models can train and improve
	Parents, caregivers and general public	▸ **Learn about current risks** of child sexual abuse online and how to best educate children on the issue in a constructive and effective way ▸ **Talk to (own) children** about using the internet, what responsible behavior online is, how to **detect** signs of grooming and how to deal with such situations ▸ If risk for own children is perceived high, **install latest safeguarding tools** on devices ▸ **Support NGOs working** on the frontlines of the problem; spread the word and make their voices heard on platforms and to governments who have the power to change policies to protect children



About the Authors

Bracket Foundation's mission is to harness the power of technology for social good by leveraging private technology solutions to tackle growing global challenges. Each year Bracket Foundation explores a theme. This year's theme is "making the internet safer for children". As a result of this endeavor, Bracket Foundation and its partners have issued a leading publication on how Artificial Intelligence (AI) can combat online sexual abuse of children. Bracket Foundation is engaged with several public sector actors which include multi-lateral organizations, NGOs and states to launch a pilot program to detect, prevent and prosecute online sexual abuse of children by leveraging AI technology. This ambitious plan requires raising awareness on the uses of AI, building trust between the public and private sector, advocacy for more government investment in AI, lobbying for more data sharing commons and changes to the legislative framework around data use in order to scale a global solution to this issue with AI as its backdrop. Bracket Foundation is the venture philanthropy arm of Bracket Capital, a leading investor in technology on the West Coast.

Bracket Capital is a global multi-asset investment manager based in Los Angeles, California. Founded in 2017, Bracket Capital develops strategies that quickly meet global market conditions by applying a systematic risk/reward methodology across the venture landscape.

Yalda Aoukar is Co-Founder/Managing Partner of Bracket Capital and Chairperson of Bracket Foundation. She is a fierce advocate of technology for good especially as it relates to solving the world's most pressing global challenges. She is a champion for women's empowerment and entrepreneurship in Venture Capital and other financial sectors, where women are traditionally underrepresented. In addition to investing in leading technology companies, she serves as an adviser to governments and policy makers on digital development in diverse fields such as Fintech, Smart City Building, Artificial Intelligence Integration and Education Technology. She sits on the board of the World Innovation Summit for Education (WISE) Accelerator providing support and guidance to international Edu-Tech start-ups. She is a member of the Fast Company Impact Council and an active member of Tomouh, an invitation-only network for young Arab leaders.

Value for Good is a consultancy specialized in the field of social impact that envisions a world in which effective action is taken to solve societal challenges. To achieve this Value for Good inspires through insights, advises through consulting and empowers through training. Value for Good serves leaders from private sector, governments, international institutions, foundations and non-profits and equips them with the knowledge and tools to make a positive and sustainable difference. For further information visit www.valueforgood.com.

Acknowledgements

This report has been produced by **Bracket Foundation** in collaboration with **Jeff Macdonald** and **Clara Péron (Value for Good)**.

The following individuals contributed with interviews and input:

Robert Beiser
- Freedom Signal (Seattle Against Slavery)

Irakli Beridze
- UNICRI Centre for Artificial Intelligence and Robotics

Adam Blackwell
- Development Services Group

Anna Borgström
- NetClean

Sarah Brown
- STOP THE TRAFFIK

Laura Clawson
- International Justice Mission

Aldo Faisal
- Imperial College London

Kevin Guo
- HIVE

Cyrus Hodes
- The AI Initiative (Future Society Initiative)

Jonas Kaiser
- Berkman Klein Center for Internet & Society

Mayank Kejriwal
- MEMEX, USC Information Sciences Institute

Dr. Unni Krishnan
- Save the Children

Dr. Florian Ostmann
- The Alan Turing Institute

Ahmed Ragab
- Harvard Kennedy School

Lloyd Richardson
- Project Arachnid

Lars Roemheld
- QuantCo

Caitlin Ryan
- Mayor's Office of Policy, Washington DC

Homayra Sellier
- Innocence en Danger

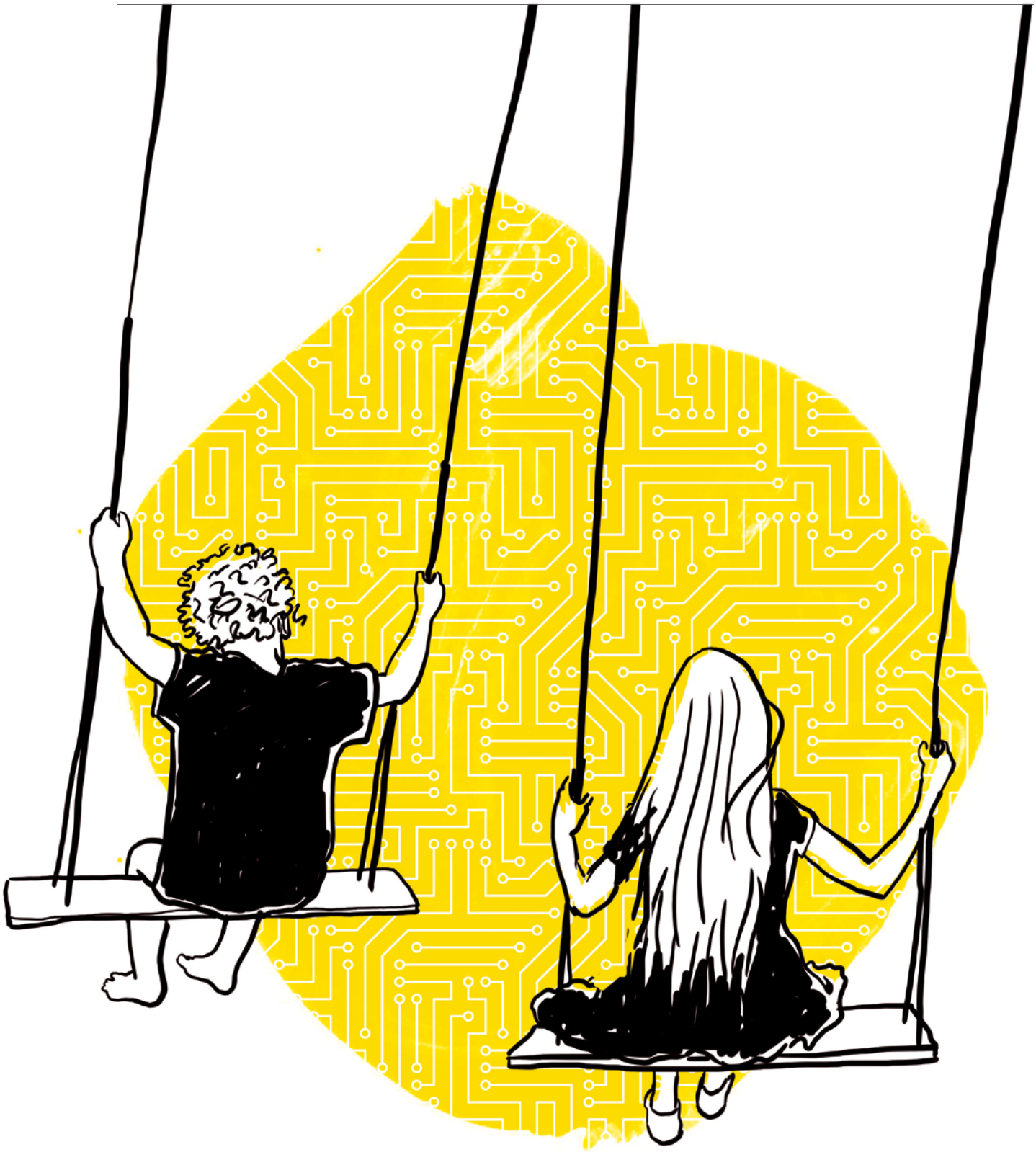
Rob Spectre
- childsafe.ai

Michael Tunks
- Internet Watch Foundation

Livia Wagner
- Global Initiative Against Transnational Organized Crime for Tech Against Trafficking

Further Reading

- Child Dignity in the Digital World. **"Child Dignity Alliance: Technical Working Group Report"**. 2018. https://www.childdignity.com/technical-working-group-report
- ECPAT International. **"Trends in Online Child Sexual Abuse Material".** In partnership with UNICEF. 2018. https://www.ecpat.org/wp-content/uploads/2018/07/ECPAT-International-Report-Trends-in-Online-Child-Sexual-Abuse-Material-2018.pdf
- Elie Bursztein. **"Rethinking the Detection of Child Sexual Abuse Imagery on the Internet".** Google, in collaboration with NCMEC and Thorn. May 2019. https://ai.google/research/pubs/pub48118/
- The Internet Watch Foundation. **"Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed Child Sexual Abuse".** In collaboration with Microsoft. May 2018. https://www.iwf.org.uk/sites/default/files/inline-files/Distribution%20of%20Captures%20of%20Live-streamed%20Child%20Sexual%20Abuse%20FINAL.pdf
- Michael C. Seto et al. **"Production and Active Trading of Child Sexual Exploitation Images Depicting Identified Victims".** Thorn, in cooperation with NCMEC. March 2018. https://www.researchgate.net/publication/325668273_Production_and_Active_Trading_of_Child_Sexual_Exploitation_Images_Depicting_Identified_Victims
- NetClean. **"NetClean Report 2017: Eight Important Insights into Child Sexual Abuse Crime"**. 2017. https://www.netclean.com/netclean-report-2017/
- UNODC. **"Study on the Effects of New Information Technologies on the Abuse and Exploitation of Children".** United Nations. May 2015. https://www.unodc.org/documents/Cybercrime/Study_on_the_Effects.pdf
- WePROTECT Global Alliance. **"Global Threat Assessment 2018: Working together to tend the sexual exploitation of children online"**. 2018. https://static1.squarespace.com/static/5630f48de4b00a75476ecf0a/t/5a85acf2f9619a497ceef04f/1518710003669/6.4159_WeProtect+GA+report+%281%29.pdf



Imprint

Authors
Bracket Foundation
9701 Wilshire Blvd, Suite 930
Beverly Hills, CA 90212
info@bracketfoundation.org

Clara Péron and Jeff Macdonald
Value for Good GmbH
Französische Str. 47, 10117 Berlin
Germany
mail@valueforgood.com

Funding Partner
Bracket Capital
9701 Wilshire Blvd, Suite 930
Beverly Hills, CA 90212
info@bracketcapital.com

Design, Layout & Illustrations
Martin Markstein, contact@derMarkstein.de