

---

# CENTER FOR DIGITAL DEMOCRACY

---



## Social Media Platform Safeguards for Whom?

Assessing Tech Industry Strategies for  
Protecting Children and Shielding Advertisers

---

Kathryn C. Montgomery  
Jeff Chester  
Katharina Kopp

NOVEMBER 2022

## Table Of Contents

<b>Executive Summary</b> .....	3
<b>Introduction</b> .....	6
<b>Content Moderation and Brand Safety Systems on Social Media Platforms</b> .....	9
Content moderation .....	9
Brand safety .....	11
<b>Platforms Under Pressure</b> .....	14
Instagram .....	14
Snapchat .....	17
TikTok .....	19
Twitch .....	22
YouTube.....	26
<b>Questions and Strategic Opportunities</b> .....	29
<b>Current U.S. Policy Proposals on Children and Technology</b> .....	32
Privacy .....	33
Safety, Design and Research .....	33
State-level Design Legislation.....	34
Reform of Section 230.....	34
<b>Building a “Digital Enabling Environment” for Young People</b> .....	35
<b>References</b> .....	37

## EXECUTIVE SUMMARY

The Facebook whistle blower revelations in 2021 have reverberated across the tech industry, amplified by a succession of congressional hearings, intense press attention, and rising pressures from abroad; all of these developments have triggered a flurry of activity from the social media industry. This report provides an overview of the major tech companies and their popular platforms, focusing on the relationships between their business operations and their child-protection policies.

The report is divided into five sections: 1) a summary of the major strategies that tech companies have developed to address problematic content, with a focus on content moderation and brand safety; 2) brief profiles of five of the leading social media platforms—Instagram, Snapchat, TikTok, Twitch, and YouTube—that attract large numbers of young people, focusing on how these companies—and the industry as a whole—have balanced the platforms’ goal of maximizing profits from young people, with the necessity of responding to regulatory, political, and public pressures over their content, advertising, and child-safety operations; 3) a discussion of the implications of our research for child advocates; 4) a brief survey of current U.S. regulatory and policy proceedings that could present opportunities for strengthening youth safeguards; and 5) a framework and strategy for ensuring that the technology industry promotes the health, safety, and wellbeing of young people.

The following are key highlights:

**In the largely unregulated tech industry, social media companies have developed a multiplicity of strategies and tactics, mobilizing the full spectrum of content-moderation systems, corporate responsibility operations, and other institutional mechanisms to demonstrate that they are capable of policing themselves.**

From industry accounts of these changes, the systems that are now in place appear to be much more expansive and robust than they were only a year or two ago, capable of tracking and taking down child sexual abuse material (CSAM) and other illegal content, but also deployed to flag and remove a growing spectrum of controversial and problematic content, such as hate speech and vaccine disinformation.

**Brand safety has become increasingly central to the functioning of the technology industry, not only providing protection for advertisers,**

**but also serving as an essential safety valve for tech companies that have found themselves under fire for their failure to police harmful content and user behavior on their platforms.**

Brand-safety systems are largely automated, and often operate on an industry-wide scale, though the major tech platforms have also developed their own internal brand-safety operations. Within the last few years, the advertising industry has established an expanding infrastructure of specialized companies, technological tools, software systems, and global consortia that now operate at the heart of the digital economy. Today’s ubiquitous brand-safety apparatus enables the most powerful marketers to have a much greater influence over tech platforms, operating a highly sophisticated surveillance system that can determine instantaneously which content can be monetized and which cannot.

**Social media platforms are now placing greater emphasis on safeguards for teens, after decades of downplaying concerns over adolescent vulnerabilities in the online environment, and resisting legislative efforts to extend privacy protections to young people over the age of 12.** The new teen-safety features are a response not only to rising controversies over social media’s harmful impacts on teen users, but also to several recent U.S. and international policy developments that are affecting the operation of global tech companies, including efforts to extend protection to older youth under the Children’s Online Privacy Protection

Act (COPPA); the EU's General Data Protection Regulation (GDPR), adopted in 2018, which identifies a person under age 18 as a minor and thus deserving of special protections; and the UK's implementation of its Age Appropriate by Design Code to protect young people online.

**Social media platforms are building automated guardrails onto the navigation experience to structure and constrain teen interactions online.** While these kinds of defaults are not new, they have become much more widespread in response to recent pressures on the entire social media industry over how its operations are negatively affecting the mental health of teens. Though they are a significant improvement, they also raise some important questions and concerns. For example, they appear to draw from the same arsenal of AI and other granular forms of surveillance, data analysis, and microtargeting that the industry has perfected for targeting teen users with marketing and advertising. To what extent are these two functions intertwined? Does the use of these safety tools expand the collection and use of data on teens?

**Some of these safety systems also raise questions about equity.** Are all teens equally affected by these tools, or might there be disparate impacts depending on gender, race, language ability, socioeconomic status, or other factors? For example, a number of the safety defaults are based on the assumption that parents and caregivers are in the best position to protect their own children from child sexual abuse material, predation and other harms. Yet this is not the case for all young people. There is research evidence that a significant percentage of child sexual

abuse images is produced by family members or close, trusted adults. Will the new safety tools protect these children or do them further harm? Children in low-income families may be at even greater risk.

**Although platforms, marketers and digital media trade groups continually claim they are concerned about young people, the industry's efforts to establish safeguards are, at best, fragmented and conflicted.** In most cases, companies have been careful not to allow any new safety systems to interfere fundamentally with their advertising practices or undermine their business models. Platforms and other online content companies are making major investments to expand the ways they can attract more young people, maximize their engagement with digital services, and generate ad and other revenues. The tech industry is also pouring hundreds of millions of dollars into building up its lobbying apparatus and contributing money to key politicians who can help them evade regulation.

**Despite the expansion of new safety systems, young people remain at risk in today's social media environment.** Numerous reports have documented continuing problems on the platforms we examined for this report. For example, in the summer of 2022, an analysis conducted by the Canadian Centre for Child Protection (C3P) found that teen boys were being subjected to "sextortion" tactics on both Snapchat and Instagram, suggesting that this ongoing problem is only getting worse, and raising questions about whether these and other social media companies are doing enough to keep children safe online.

**Advertisers are a powerful force in the entire digital media ecosystem, yet their increasingly influential role in content decisions has remained largely absent from the public debates over children and the internet.** The industry has established an expanding set of institutions, policies and technologies that are shaping the operations of the leading platforms and publishers, continually fine-tuning these systems to minimize risks while maximizing monetization. Can the affordances of the brand-safety complex, including ad targeting and analysis, be used to assess CSAM and other harmful content more effectively? How well do the current mechanisms work? The child-safety field needs to have an assessment of the investment and technology-deployment decisions platforms are making in AI and other content-recognition technologies. How are they being implemented in the emerging metaverse, and to what effect? There are also a number of fault lines and conflicts within the industry that could create opportunities for child advocates to press for better protections on behalf of young people.

**Even with the tech industry's recent moves toward greater transparency, most of the operations inside these social media companies remain hidden from public view, leaving many questions about how the various safety protocols and teen-friendly policies actually function.** For example, how well and in what ways do brand safety, content moderation, and by-design navigation tools work hand in hand? How much control do advertisers have over content? What kinds of responsibilities are being assumed by other participants in the online social media ecosystem? A key question



is whether the platform and online industry commitments to making the internet safer for advertisers is a more comprehensive and economically supported effort than what the companies do with content moderation and other efforts related to the safety of youth. Does brand safety overshadow child-safety efforts, given that the former's role is to enable successful ongoing revenue generation from advertisers? Are current brand-safety systems effectively preventing the monetization of CSAM-related sites?

**The most effective way to ensure greater accountability and more meaningful transparency by the social media industry is through stronger public policies.** We are now witnessing an unprecedented political environment, in which policymakers on both sides of the aisle are calling for government safeguards to protect children. At this point, there is strong interest in Congress for policies that would address three major areas related to the tech industry's relationship to young people: updating the Children's Online Privacy Protection Act (COPPA); instituting new design requirements to govern how young people interact with social media and other digital platforms; and strengthening safeguards against CSAM and other sexual abuse practices, which includes reforming the immunity protections that tech companies now have under Section 230 of the Telecommunications Act. While versions of these bills have been introduced in prior Congressional sessions, some elements are new and others have been retooled and updated in light of recent controversies in the tech industry.

**Tech policies in the U.S. have traditionally followed a narrow, piecemeal approach to addressing children's needs in the online environment, providing limited safeguards for only the youngest children, and failing to take into account the holistic nature of young peoples' engagement with the digital media environment.** Thus there is a need for more comprehensive and integrated policies that combine privacy protections, advertising and commercial surveillance safeguards, with child safety and CSAM protection. This will require a strategic effort that brings together the diverse constituencies working on behalf of youth in the online media. Because the impacts of digital technologies on children are so widespread, efforts should also be made to broaden the coalition of organizations that have traditionally fought for children's interests in the digital media to include groups representing the environment, civil rights, health, education, and other key stakeholder communities.

**As part of a long-term strategy, all stakeholders concerned about young people should work together to identify the kinds of outcomes we wish our children to experience from the online media culture.** One approach would be to draw from the field of positive youth development, which calls for an "enabling environment" for young people, one that "supports their assets, agency, access to services, and opportunities, and strengthens their ability to avoid risks and to stay safe, secure, and be protected and live without fear of violence or retribution," and that "encourages and recognizes youth,

while promoting their social and emotional competence to thrive." Though this concept was not devised specifically for digital technologies, its principles and goals could become key pillars in a framework for a "digital enabling environment" for young people. The framework should also be rooted in the recommendations adopted by the United Nation's Convention on the Rights of the Child (UNCRC), declaring that the rights of "every child" around the world must be "respected, protected and fulfilled in the digital environment." In addition to addressing the needs of the individual child, we must also take into account the unique experiences of the United States as a nation, ensuring that digital technologies produce equitable outcomes and enable equal opportunity for specific groups of young people, including those from marginalized socio-economic and ethnic backgrounds.

## INTRODUCTION

In September 2021, former Facebook employee Frances Haugen appeared before Congress to testify about the company's internal research on how its Instagram photo-sharing app negatively affected the mental health of teens. The explosive revelations cast a rare spotlight on the internal operations of the powerful social media platform. A series of articles based on the documents that Ms. Haugen released to the news media revealed how Facebook had conducted a relentless behind-the-scenes campaign to push its services further into children's lives, amassing an expanding trove of internal research to probe the deep emotional states of its youth users, including those under the age of thirteen, who are not officially allowed on Instagram.<sup>1</sup> In widely publicized hearings, members of Congress from both sides of the aisle expressed outrage at Facebook's actions, and leveled criticism more broadly at the entire technology industry. Senator Richard Blumenthal (D-Conn.) charged Instagram and other social media with creating "addictive" platforms that are exacerbating the mental health crisis among young people through powerful algorithms that "exploit and profit from children's insecurities and anxieties."<sup>2</sup> Representative Ken Buck (R-Colo.) accused Facebook of lying, and declared that "Big Tech has become the new Big Tobacco."<sup>3</sup>

Facebook subsequently changed its corporate name to Meta, ostensibly to reflect more accurately the future of digital media, which is evolving into an all-encompassing "metaverse," described as "the next evolution of social connection," dominated by

virtual reality, augmented reality, and other immersive technologies.<sup>4</sup> But the brand shift will likely not release Facebook/Meta from the controversies that have surrounded the company since it was launched in 2004. The most recent incident is emblematic

of an ongoing, inherent conflict that will likely persist as the digital media culture evolves.

Since the beginning of the internet era, children and youth have been a key demographic target for the tech industry, valued for their role as early adopters of technology, and highly sought by advertisers because of their lucrative spending power.<sup>5</sup> The recent revelations from the "Facebook Files" illustrate just how valuable young people are to online platforms. As reported in October 2021, the loss of teens by Facebook in recent years was considered an "existential threat" because of the importance of young people to user growth. Consequently, in 2018, Facebook "earmarked almost its entire global annual marketing budget to targeting teenagers, largely through digital ads." The document revealed that Facebook was especially concerned about reaching 13-15 year-olds. Facebook's spending to promote Instagram was said to be \$390 million a year, most of it designated to attracting teens. Teenagers are key to social media spending and help make up the \$150 billion in "collective buying power" for the "Gen Z" cohort, according to eMarketer.<sup>6</sup>

Tech companies specifically design many of their platforms and services to tap into the psychological, social, and emotional needs and experiences of adolescents, who are considered important trend setters for influencing how the broader public will engage with digital media.<sup>7</sup> As children and teens have eagerly embraced a growing number of social media,

gaming platforms, and mobile apps, commercial imperatives have shaped the structures and operations of their digital experiences. But many of the same technological features that make digital platforms both attractive to young people and lucrative for advertisers are also jeopardizing their privacy, targeting them with manipulative marketing techniques, and subjecting them to inappropriate content, cyberbullying, sexual predation, and other harms. These features include, for example, user-generated content, livestreaming video, and algorithms that demand constant attention, foster engagement, and promote sharing.

Among the risks that young people encounter when they are online is exposure to child sexual abuse material (CSAM). Though much CSAM activity takes place on the “dark web,” a good deal of it occurs on mainstream social media platforms.<sup>8</sup> While all of the leading platforms are required by federal law to report child sexual abuse material to the National Center for Missing and Exploited Children’s CyberTipline, the problem has only gotten worse, expanding exponentially over the last two decades.<sup>9</sup> As online child sexual exploitation continues to rise at alarming rates, purveyors of CSAM are tapping into a growing number of commercial digital media platforms and applications that attract large numbers of youth. Among them are Amazon’s gaming site, Twitch; Facebook’s Messenger, WhatsApp and Live services; Google-owned YouTube; and video-sharing app TikTok.<sup>10</sup>

Except for being subject to a handful of legal requirements, the technology industry in the U.S. has been allowed to evolve and grow without any

***The growing power of global Big Tech platforms creates an urgent need for comprehensive, coordinated, international efforts to promote greater corporate accountability and stronger government policies on behalf of children, and to ensure that safeguards are built into the foundation of emerging platforms and services.***

comprehensive policies to guide its development. Although privacy and data-protection laws—including both the General Data Protection Regulation (GDPR) in the EU and the Children’s Online Privacy Protection Act (COPPA) in the U.S.—have placed some limits on advertising and data practices, including those that target young children, these policies have done little to slow the growth of today’s expansive and sophisticated digital marketing operations.<sup>11</sup> The pandemic, moreover, triggered a dramatic increase in young people’s engagement with digital media, driving up profits for social media, gaming platforms, streaming services, and other tech companies, while exacerbating the risks for youth. Over the years, individual companies have introduced a series of corporate measures designed to assure parents, advertisers, investors, and regulators that children are safe when engaging with their products and services. However, all of these self-regulatory initiatives have been piecemeal, sometimes sparked by controversy, and frequently designed to offer only the most limited degree of protection, deflecting criticism while ensuring that commercial operations will proceed unimpeded.

The growing power of global Big Tech platforms creates an urgent need for comprehensive, coordinated, international efforts to promote greater corporate accountability and stronger government policies on behalf of children, and to ensure that safeguards are built into the foundation of emerging platforms and services. These efforts must be grounded in a deep understanding of how the technology industry, its commercial practices, and its technical operations are shaping young people’s interactions with the digital environment. Protecting children from online sexual exploitation should also be part of a broader agenda for young people that encompasses not only safety, but also privacy, fairness, equity, and other critical rights. Advocates need to take advantage of a number of recent developments in both the technology industry and the public policy arena, seizing the opportunities and addressing the challenges.

The Facebook whistleblower controversy erupted amid a growing tech-accountability movement, with policymakers, the news media, and civil society organizations taking aim at the powerful technology giants over

a growing spectrum of issues—from enabling widespread distribution of disinformation campaigns and hate speech, to inadequate protections against sexual predation and exploitation, to anti-competitive behaviors in the marketplace.<sup>12</sup> U.S. policymakers in Congress and at the Federal Trade Commission (FTC) are placing greater scrutiny on the operations of these companies, introducing legislation, proposing regulatory policies, and holding hearings.<sup>13</sup> A number of bills have been introduced in Congress, including several that would specifically protect young people.<sup>14</sup> Some state regulators are passing their own laws.<sup>15</sup> And policymakers in Europe and the UK are instituting new regulations on the design and operations of social and other digital media platforms, which are reverberating globally.<sup>16</sup>

We have been tracking the growth of the digital marketplace for more than 25 years, closely following the incredible expansion of the technology industry and its increasingly sophisticated advertising, data-collection, and marketing apparatus.<sup>17</sup> We have now also extended our analysis of the digital marketplace to include how the tech industry has addressed CSAM and other child-safety issues. In the following pages, we will summarize what we have learned about the major tech companies and their popular platforms, focusing on the relationships between their business operations and their child-protection policies. We have undertaken this recent examination within the context of the larger set of developments that are influencing the tech industry, and through the lens of our longstanding expertise in digital practices related to children and youth.

In response to the increasing pressure over their failure to address a litany of problematic content, social media companies have promised to make their content-moderation and safety practices less obscure, opening “transparency centers” and publishing a series of transparency reports. While we have been able to learn something from these reports, the documents are often selective in what they choose to reveal, and are often strategically framed with the goal of assuring policymakers that there is no need for government intervention or regulation. Fortunately, the tech-accountability movement has resulted in greater press coverage of the industry’s otherwise hidden operations. Civil society organizations are also playing a role in closely monitoring tech company practices. In order to develop a clearer and more accurate portrait of these companies, we have relied on news reports, trade articles, and industry analyses. In addition, we have drawn from our own experiences interacting with some of these corporations and observing their operations at close range during meetings and negotiations over youth policies and practices. Though our information may be limited, we are able to identify some patterns, themes, and issues that advocates can use in their ongoing efforts to ensure that adequate safeguards are in place to protect young people.

This report is divided into five sections:

1) We first provide an overview of the major strategies that tech companies and their social media platforms have developed to address problematic content, with a focus on content moderation and brand safety, both of which have become central tools in handling issues related to child sexual

abuse and youth protection. 2) We then present brief profiles of five of the leading social media platforms that attract large numbers of young people: Meta-owned Instagram; Snapchat; TikTok; Twitch; and YouTube. Our focus is on how these companies—and the industry as a whole—have balanced the platforms’ goal of maximizing profits from young people, with the necessity of responding to regulatory, political, and public pressures over their content, advertising, and child-safety operations. 3) We discuss the implications of our research for child advocates, and suggest some questions for further investigation, along with interventions for holding the technology industry more accountable to children and teens. 4) Our next section briefly covers current U.S. regulatory and policy proceedings that could present opportunities for strengthening youth safeguards. 5) Finally, we suggest a framework and strategy for ensuring that the technology industry promotes the health, safety, and wellbeing of young people.<sup>18</sup>



## CONTENT MODERATION AND BRAND SAFETY SYSTEMS ON SOCIAL MEDIA PLATFORMS

Content moderation and brand safety are two of the major strategies that the tech industry employs to deflect external criticism, protect advertisers, and pre-empt government regulation. While each of these internal functions has developed separately, they have become increasingly integrated, and have expanded in response to rising pressures on the industry in recent years. Today, they are part of a growing, industry-wide, largely automated, tech-safety apparatus, which operates throughout the social media ecosystem.

### CONTENT MODERATION

Under Section 230 of the Telecommunications Act of 1996, social media companies, also known as “information intermediaries,” cannot be held liable for the content posted by their users. At the same time, they are not simply neutral conduits, but do maintain the power to police user-generated content that violates their policies. “If an intermediary does police what its users say or do, it does not lose its safe harbor protection by doing so,” explains scholar Tarleton Gillespie, whose book, *Custodians of the Internet*, is the most comprehensive study to date of content moderation in the technology industry.<sup>19</sup> As he explains, the legislative language was carefully crafted “so that the safe harbor would not create legal jeopardy for intermediaries that chose

to moderate in good faith, by making them more liable for it than if they had simply turned a blind eye.”<sup>20</sup>

Content moderation has evolved from the early formative period of the internet, when volunteers in online chat rooms assumed responsibility for monitoring conversations in order to report any content considered “offensive.” Today’s operations are large, highly complex, sophisticated enterprises, employing not only full-time “Trust and Safety” professionals, but also “temporary crowdworkers, outsourced review teams, legal and expert consultants, community

managers, flaggers, admins, mods, superflaggers, nonprofits, activist organizations and the entire user population.”<sup>21</sup> Many of these workers are hired as independent contractors by third-party companies, such as Amazon’s Mechanical Turk, Upwork, Accenture, and TaskUs.<sup>22</sup> The dramatic expansion of content-moderation systems is directly tied to the social media industry’s business model. “On Facebook alone,” explained a recent article by NYU’s Stern Center, “more than three million items are reported on a daily basis by users and artificial intelligence screening systems as potentially warranting removal. This degree of volume didn’t happen by accident; it stems directly from Facebook’s business strategy of relentlessly pursuing user growth in an effort to please investors and the advertisers that are the company’s paying customers.”<sup>23</sup> What began as a set of one-page moderation policies in the early days of the social media network has evolved into a 15,000-word document that “remains the basis of the company’s publicly available Community Standards, which have been amended many times over the years.”<sup>24</sup> During the huge controversy triggered by the Cambridge Analytica scandal in 2016, Facebook increased the number of its content moderators by two-thirds, although most of them

***The dramatic expansion of content -moderation systems is directly tied to the social media industry’s business model.***

were outsourced. Today, its content moderation takes place at 20 sites worldwide.

Though individual companies operate independently of one another, their approaches to content-moderation have been quite similar. Policies and practices are ostensibly based on terms of service, community guidelines, and other public materials. However, “there are real differences between the rules these companies post and the decisions they end up making, case by case, in practice,” explains Gillespie. Rather than engaging in the kind of pre-publication editorial gatekeeping that a broadcaster or publisher does, tech platforms adopted a “publish-then-filter” approach, where “everything, no matter how reprehensive or illegal, can be posted to these platforms and will be available until it is noticed and removed.”<sup>25</sup> The process relies extensively on “community flagging,” where users bear much of the responsibility for identifying and reporting offensive content and behaviors. This system is far from perfect and can be particularly burdensome for individuals reporting sexual abuse. For example, research conducted by the Canadian Centre for Child Protection found that on some of the most popular platforms, it was easier for users to report issues like copyright infringement than it was to

report child sexual abuse material.<sup>26</sup>

In recent years, tech companies have been deploying machine learning and artificial intelligence to flag problematic content on their platforms, though experts note that these systems are still limited. These mechanisms can be useful for identifying and removing CSAM and other illegal material.<sup>27</sup> Photo DNA, which Microsoft first introduced in 2009 and made available free to social media platforms, uses a technique called “hashing,” where a digital image can be converted into a readable code, and then matched to other images.<sup>28</sup> The unique hashes of all known child pornography images are maintained by the National Center for Missing and Exploited Children (NCMEC). If a new image is a match, the platform can remove it. “Every single photo you post to Instagram, every image you pin on Pinterest, every snap on Snapchat, is quickly and automatically compared to the NCMEC database to make sure it is not child pornography,” explains Gillespie. As Michael Salter and Elly Hanson write, the introduction of Photo DNA was a “major turning point in the fight against CSAM, and the determining factor in the dramatic numbers of CSAM currently being notified to U.S. authorities.” The authors point out, however, that it took as long as five years to implement Photo DNA, and there

has been a “lack of significant industry investment in the further development and deployment of the technology.”<sup>29</sup> Moreover, as advanced as it is, this technology is unable to identify new images of CSAM or to detect video files, even though video CSAM is more common today than photos.<sup>30</sup>

AI and machine learning are also being used to address a much broader range of problematic content, significantly altering many of the content-moderation operations within the largest social media platforms. The coronavirus pandemic has accelerated a trend that was already underway in the industry, triggering an increased reliance on automated systems, many of which have remained in place. For example, Meta reported that it turned to its AI systems when the company temporarily sent its content reviewers home, and that it now relies on AI to “help prioritize content for review,” which it claims has improved its “proactive detection” rate, enabling it to take action on 25.7 million pieces of “child sexual exploitation” content on Facebook and 1.4 million on Instagram.<sup>31</sup> TikTok also relies on “automated defenses to detect, block and remove” content.<sup>32</sup> These statistics are part of a 2021 report to the Global Alliance for Responsible Media (GARM), an industry-wide initiative that is a leading force in the brand-safety movement.

The rise of brand safety—which relies extensively on Big Data advertising, targeting, and measurement operations—is another factor that has helped transform today’s content-moderation systems. Today, content moderation and brand safety are working more and more in tandem with each other.

***AI and machine learning are also being used to address a much broader range of problematic content, significantly altering many of the content-moderation operations within the largest social media platforms.***

## BRAND SAFETY

Brand safety is designed to shield advertisers from being associated with a growing number of contentious and controversial issues. The list of forbidden topics has become longer and more granular in recent years, encompassing illegal content, hate speech, violence and sexually explicit material. Brand-safety systems are largely automated, and often operate on an industry-wide scale, though the major tech platforms have also established their own internal brand-safety operations. Like content moderation, brand safety has become increasingly central to the functioning of the technology industry, not only providing protection for advertisers, but also serving as an essential safety valve for tech companies that have found themselves under fire for their failure to police harmful content and user behavior on their platforms. Within the last few years, the advertising industry has established an expanding infrastructure of specialized companies, technological tools, software systems, and global consortia that now operate at the heart of the digital economy. Today's ubiquitous brand-safety apparatus enables the most powerful marketers to have a much greater influence over tech platforms, operating a highly sophisticated surveillance system that can determine instantaneously which content is monetized and which is not.<sup>33</sup>

Brand safety's origins can be traced to the introduction of programmatic advertising. Through this system—introduced about ten years ago as “real-time bidding”—advertisers don't buy time, as they have done in more traditional media such as television or radio. Nor do they place their

***Brand safety is designed to shield advertisers from being associated with a growing number of contentious and controversial issues. The list of forbidden topics has become longer and more granular in recent years, encompassing illegal content, hate speech, violence and sexually explicit material.***

ads in specific programs or content. Instead they tap into a sophisticated automated software system that can find and target individual users wherever they go, online and off. This online ad-buying process has also allowed all manner of content to get financial support from ads—even when it involved misinformation, hate speech, threats to young people or other harmful activities. When the news media began to report in 2017 how ads from well-known brands were found adjacent to hate speech, child exploitation, and other objectionable content on YouTube and Facebook, the online ad industry came up with a set of mechanisms for protecting advertisers from association with such content.<sup>34</sup> These operations were also designed to address ad fraud, a major loss of revenue for marketers, as well as to ensure that people are actually able to view ads (a key industry metric called “viewability”).<sup>35</sup>

In 2019, the World Federation of Advertisers (WFA), which includes some of the most powerful brands and ad agencies in the world, helped launch the Global Alliance for Responsible Media (GARM), which has taken on the task of establishing

international standards for addressing online safety within the advertising industry.<sup>36</sup> GARM has worked with its members to develop a classification system for what it calls “non-brand-safe and suitable content,” enlisting the cooperation of major technology companies, including Facebook (Meta), YouTube and Twitter, to support the framework and help coordinate its implementation.<sup>37</sup> The classification system includes granular descriptions of eleven different content categories considered inappropriate for any advertising. Among them are “adult and explicit sexual content,” “the illegal sale, distribution, and consumption of child pornography,” the “explicit or gratuitous depiction of sexual acts, and/or display of genitals, real or animated,” hate speech, “graphic promotion, advocacy and depiction of willful harm and actual unlawful criminal activity,” and “harassment or bullying of individuals and groups.”<sup>38</sup>

However, this classification system—which places large swaths of content off-limits—is not set in stone. Since it was initially developed, it has evolved in response to advertisers who sought greater flexibility about where they could invest their dollars

online. The amended “Brand Safety Floor and Suitability Framework,” released in 2020, offers advertisers a more nuanced set of categories, high, medium, or low risk, for identifying which content to monetize. For example, high risk would involve “suggestive sexual situations requiring adult supervision/approval or warnings,” as well as “full or liberal nudity.” Medium risk is defined as the “dramatic depiction of sexual acts or sexuality issues presented in the context of entertainment,” as well as “artistic nudity.” Low risk refers to “educational, informative, scientific treatment of sexual subjects or sexual relationships or sexuality.”<sup>39</sup>

Advertisers contract with third-party firms that independently measure marketing campaigns related to brand safety and a set of related advertising performance measures on the major tech platforms.<sup>40</sup> The two industry leaders are Integral Ad Science (IAS) and DoubleVerify (DV). Both are fully incorporated into the commercial online ecosystem, offering a sophisticated set of technological tools for implementing brand-safety strategies on a continuous basis across all the major platforms.<sup>41</sup> Both IAS and DV continually “scan” the content available from the leading platforms, as well as other online media. They each deploy cognitive-semantic, machine-learning and AI technologies, available in dozens of languages, to assess online content at the page level.<sup>42</sup> The IAS “data-ingestion” system includes a “web infrastructure that fetches hundreds of millions of web pages per day [and] edge measurement servers that collect thousands of data points per transaction...” in order to process “over 100 billion web transactions per day” for “trillions of data events per

## ***A number of brand-safety companies specifically offer services aimed at children’s safety.***

month.”<sup>43</sup> Competitor DV notes that it is “integrated across the entire digital advertising ecosystem” and can evaluate “200 billion transactions” a day.<sup>44</sup>

Third-party brand-safety companies have identified dozens of categories of content that can be avoided when delivering ads and marketing. Sites are assessed and “scored” and also sorted into “whitelists” and “blacklists” that can control where and how any ads are placed.<sup>45</sup> Clients of DV can “set profiles and adjust controls” for their online advertising efforts, including “settings such as inclusion/exclusion lists, override lists, custom keyword avoidance, inappropriate content avoidance categories” as well as “over 200 analytics and reporting metrics” and other data (including geolocation-related information).<sup>46</sup> Third-party and independent brand-safety services are worldwide in scope; IAS, for example, says it serves 111 countries.

A number of brand-safety companies specifically offer services aimed at children’s safety. For example, OpenSlate, which works with YouTube and Facebook, offers “brand suitability models” to protect advertisers “from running against channels that produce content intended for children 12 years and under.”<sup>47</sup> Pixilate, a brand-safety company that works with Amazon, among others, offers a “child safety risk” analysis for ad-supported online content, including apps, streaming video and online content available on websites and via mobile devices.

Among the categories it identifies are “adult” and violent content, “brand safety risks,” and whether there are concerns related to potential violations of the Children’s Online Privacy Protection Act (COPPA).<sup>48</sup>

But the elaborate protective machinery for shielding advertisers from exposure to controversy still has a number of problems to tackle.<sup>49</sup> For example, Mozilla issued a report in July 2021 that documented repeated problems with YouTube and its recommendation algorithm, including “wildly inappropriate” children’s cartoons.<sup>50</sup> In February 2022, The Times of London reported how several well-known advertisers were being placed on “misinformation and conspiracy videos” on YouTube.<sup>51</sup> These problems have prompted some in the advertising industry to call for greater rigor within the brand-safety system. According to a December 2021 survey of leading global advertisers, these companies see “the need for a robust programme of media governance and new principles for media buying” over the next 12 months. In addition, a majority of companies surveyed said they “believe the media industry would benefit from stronger external regulation.”<sup>52</sup>

The industry has established its own self-regulatory evaluation mechanisms for brand safety, although independent observers have raised questions about their adequacy. For example, GARM conducts “Aggregated Measurement



Reports” that examine the “metrics and measures” used by the platform members for their transparency reports.<sup>53</sup> These analyses focus on such measures as how much harmful content is seen by consumers, how “safe is the platform for advertisers,” and “how effective is the platform in enforcing its safety policies.” In its report released in April 2021, GARM reviewed submissions from Facebook, YouTube, Snapchat, TikTok, Twitter and Pinterest concerning their compliance with the organization’s framework. The report said that the amount of content removed was more than 5.3 billion pieces, primarily from “GARM content categories Spam, Sexual Content, and Hate Speech and Acts of Aggression.” It concluded that both YouTube and Facebook had made significant advances in addressing hate speech and acts of aggression content. It also highlighted that the use of “automation and machine learning accelerated” in the context of the initial stage of the pandemic.<sup>54</sup> But as Ad Age commented, “the services still have plenty of work to do to tackle brand safety issues.” Among the report’s shortcomings, it noted, was the failure of Instagram to “divulge data on sexual content and nudity, hate speech

or spam,” although Google YouTube and Facebook “provided more data” than the other platforms covered in the report.<sup>55</sup>

As the tech industry continues to face threats of regulation, it has made a number of moves to address criticisms and bolster the legitimacy of its content-moderation and brand-safety operations. For example, in February 2021, a number of major tech companies, including Google, Meta (Facebook), Microsoft, Pinterest and Twitter, launched an industry-wide organization called the Digital Trust & Safety Partnership (DTSP), “a new initiative focused on promoting a safer and more trustworthy internet.” The organization’s “Safe Framework” assessment plan reflects “35 best practices” that their members have adopted to “keep users of digital services safe from abuse.” The group

opened a “public consultation” that ended on February 15, 2022, explaining that it especially wanted comments from civil society representatives, including safety and consumer NGOs.<sup>56</sup> Tech companies also partner with outside groups -- including Anti-Defamation League, National Domestic Violence Hotline, ConnectSafely, Family Online Safety Institute (FOSI), GLAAD, the Center for Digital Technology (CDT) and Internet Watch Foundation -- and many social media platforms have established their own advisory councils, such as Twitter’s Trust and Safety Council, Facebook’s “supreme court,” and Twitch’s Safety Advisory Council.<sup>57</sup>

As we report in the next section, even with these elaborate safety systems in place, questions still remain about how effectively the social media ecosystem is protecting young people.

***As the tech industry continues to face threats of regulation, it has made a number of moves to address criticisms and bolster the legitimacy of its content-moderation and brand-safety operations.***

## PLATFORMS UNDER PRESSURE

Below we present brief profiles of five major social media platforms popular with young people, identifying and highlighting some of the recent trends within the tech industry, with a focus on corporate initiatives for addressing concerns over children and teens. The Facebook whistleblower revelations have reverberated across the digital landscape, amplified by a succession of congressional hearings and intense press attention. The industry is also feeling pressures from abroad, as the EU and the UK institute new policies for regulating how online services interact with young people.<sup>58</sup> All of these developments have triggered a flurry of activity from these companies. As Insider Intelligence reported, “social platforms have been racing to introduce safety features for underage users and to get the word out to regulators and advertisers that their products are safe for children.”<sup>59</sup> As specific companies or platforms have been singled out in congressional investigations, targeted with consumer complaints, or embroiled in controversies, they have instituted new and improved mechanisms aimed at addressing the particular concerns, sometimes generating mixed responses from the public, policy makers, and child advocates. Among the new safeguards are default operational functions, based on a user’s age, that are built directly into the online navigation experience. As pressures have intensified, some companies have made additional tweaks to their new policies, resulting in a dizzying array of online tools, pop-ups, and other user options.

## INSTAGRAM



When it was launched in 2010, Instagram enjoyed a rapid adoption rate, acquiring one million users within the first two months. Facebook bought Instagram in 2012, promising regulators and the public that it would keep the app separate and independent from its main platform operations. However, within a few years, CEO Mark Zuckerberg had reneged on that promise, announcing in 2016 that the company was going to integrate its message services on Facebook, Instagram, and WhatsApp in order to make them more “interoperable.” But as Sheera Frenkel and Cecilia Kang explain in their recent book, *An Ugly Truth: Inside Facebook’s Battle for Domination*, this move was not just a technical fix; rather, it was clearly designed to facilitate the growth of advertising on Facebook.<sup>60</sup>

Because of its success in attracting advertisers who want to reach young people, Instagram has become particularly valuable to its parent company, Meta.<sup>61</sup> It is now one of the most popular and profitable social

media platforms in the tech industry, with teens among its most avid users.<sup>62</sup> In the U.S., 57 percent of teens use Instagram, and a recent survey placed it as the third-ranking site for teens (trailing TikTok and Snapchat).<sup>63</sup> It offers young people an easy, accessible, and fun way to share photos and videos, altering the images with filters and other tools, and tagging them by location. One of the app's most popular features is Instagram Stories, which enables users to create their own personalized slide show, capturing every moment of their day and rendering it into a compelling narrative that is designed to remain visible for only 24 hours.<sup>64</sup>

Meta's content moderation and safety operations encompass both Facebook and Instagram. The company publishes community standards for a broad spectrum of issues, including "safety, objectionable content, violence and criminal behavior, respecting intellectual property, integrity and authenticity and content-related requests and decisions." "Child Sexual Exploitation, Abuse and Nudity" are classified under safety, while "sexual solicitation, violent and graphic content and adult nudity and sexual activity" are covered under objectionable content.<sup>65</sup> In its transparency reports, Meta explains how its AI and machine-learning systems are able to identify such content, remove images and report incidents to NCMEC. It has also developed a number of "warning screens" and similar notifications that automatically appear when a user encounters problematic content.

But even before the release of the whistleblower documents, regulators in Europe, the UK, and the U.S. were

raising concerns about Instagram's harms to young people. For example, in 2019, the leading charity in the UK, the National Society for Prevention of Cruelty to Children (NSPCC), released a study showing a "200% rise in recorded instances in the use of Instagram to target and abuse children," calling Instagram the leading platform for child grooming in the UK.<sup>66</sup> Instagram has also been implicated in teen suicides in the UK. A survey by Ditch the Label, an anti-bullying nonprofit, found that 42 percent of people between the ages of 12 and 20 had experienced cyberbullying on Instagram, the highest percentage of any platform measured.<sup>67</sup> These findings have prompted UK authorities to pressure Instagram and its parent company to offer new safety features to protect young people.<sup>68</sup> The U.S. child-protection organization, Thorn, published a study in May 2021 that found Instagram to be highly prone to risks for minors, noting that more than 25 percent of respondents to its survey reported experiencing sexual enticement, bullying, and other potentially harmful experiences while on the popular app.<sup>69</sup>

Facebook has established its own brand-safety system, in response to a succession of controversies that have put advertisers in proximity to problematic content. For example, in 2020, Facebook was the focus of an ad boycott over its role in promoting racism, forcing it to make a number of changes in its content and advertising operations.<sup>70</sup> Its "brand safety center" offers a variety of "filters" for brands to choose in placing their ads. There are also controls to stop ads from appearing in Live Stream videos.<sup>71</sup> In 2021, bowing to further pressure from

advertisers, the company announced a new "topic-exclusions" capability, which looks for "accounts [users] that are most likely to have engaged with posts that fall under the topics like crime, tragedy, news and politics," and then restricts the kinds of ads that can be shown to those users.<sup>72</sup> Facebook provides brand-safety company OpenSlate access to "metadata and video metrics" so it can generate what's called a "SlateScore," which "grades video content quality for engagement, consistency, momentum, reach and brand safety."<sup>73</sup>

Today, Instagram finds itself at a critical crossroads. While continuing to garner significant profits as a key vehicle for connecting advertisers to the lucrative teen demographic, it faces a flurry of criticism from multiple sources. It is considered a key to Facebook's strategy for guaranteeing a steady flow of young users across the company's social media platforms. As the research released by former Facebook employee Frances Haugen revealed, Facebook has been directing considerable resources to exploring how to target an even younger market. But any move into the under-13 demographic—which is the only age group subject to current U.S. privacy regulations—also carries with it considerable risk. This has required careful political maneuvering as well as strategic communication and public outreach efforts. For the past several years, Facebook (and Instagram) have been building what it calls a "youth pillar" within its Community Product Group, designed to help it navigate these tricky waters of commercial expansion and management of criticism.<sup>74</sup>

In March 2021, the company announced that it was updating its

safety and privacy safeguards for users between the ages of 12 and 17, and instituting AI and machine-learning software tools to detect more aggressively under-age children trying to access the platform. For example, a teen user who chose to have a public account would be automatically sent a message touting the benefits of having a private account.<sup>75</sup> And teens who were interacting with adults would be warned in cases where the app detected that an adult engaged in suspicious behavior, such as sending a large number of friend requests to users under 18. In partnership with two nonprofit organizations, ConnectSafely and The Child Mind Institute (both of which receive funding or support from the company), Instagram issued an updated parents' guide revealing a new set of safety features.<sup>76</sup>

However, when Instagram announced that same month that it was planning to launch an Instagram Kids app, child advocates and regulators mounted a strong public protest, joined by attorneys general in 44 states, who wrote to urge the company not to move forward with its plans, citing the harmful effects of social media on children, as well as Facebook and Instagram's poor record in safeguarding young people.<sup>77</sup> Lawmakers summoned Facebook CEO Mark Zuckerberg to appear with other social media company executives to explain how they were going to ensure child safety on their platforms. Even as Facebook and Instagram continued to make further safety announcements to assuage the concerns of policymakers and parents, their efforts were eclipsed by the explosive coverage and congressional hearings triggered by the release of the whistleblower's documents. In September 2021, just one day before its

CEO was to appear before the Senate Commerce Committee, Facebook announced that it was pausing the development of the new kids app to "work with parents, experts, policymakers, and regulators."<sup>78</sup>

Instagram's recent youth safeguards are emblematic of overall trends in the social media industry, as more and more platforms popular with young people institute a spectrum of automatic, default mechanisms designed to structure the online experiences of youth users. For example, Facebook and Instagram have now introduced default settings that make the accounts of all users under 16 automatically private, and block advertisers from directing interest-based ads at them. Through the use of AI technology, the company also instituted mechanisms for identifying and blocking adult users engaged in "potentially suspicious behavior" from interacting with teens' accounts, and restricting other adult users from contacting teenagers who weren't already following them. Teens will also be notified if anyone is engaging in suspicious behavior, and will be provided with tools for blocking and reporting those individuals.<sup>79</sup> In 2021, the company launched a new feature to address concerns over internet addiction, by allowing young users to create reminders for themselves every 10, 20 or 30 minutes, prompting them to "take a break" from using the app.<sup>80</sup> In December of that year, right before Adam Mosseri, the head of Instagram, was scheduled to testify before a Senate committee on the controversial mental health research regarding teens, the company made yet another announcement, this time releasing a set of new tools to give parents and guardians greater control over their

teens' social media use. These controls would enable a parent or guardian to determine how much time their teen was spending on the platform and to set limits on screen use.<sup>81</sup> In June 2022, Instagram announced another set of parental "supervision tools" covering its virtual reality headsets, along with "nudges and reminders" encouraging teens to take periodic breaks while engaging with the platform.<sup>82</sup>

Like its teen-safety policies, the changes that Facebook announced for its teen ad-targeting policies, on its main platform and on Instagram, are an example of how the company carefully crafts its safeguards to assuage criticism, while ensuring that its operations are not significantly altered. Facebook framed its recent advertising policy changes for young people as a "more precautionary approach" in response to ongoing concerns by youth advocates about adolescent vulnerability to many of the marketing and data practices on today's social media platforms. However, as Digiday explained, "agency execs who eat, sleep and breathe Facebook advertising say the company's automated ad targeting algorithm could continue doing a good—possibly even better—job of targeting ads to teens whom the platform deems likely to respond to their ads."<sup>83</sup> While the company will not allow advertisers to target specific individuals, or to deploy what is known as "lookalike audience targeting" to reach young people under the age of 18, such limits are not likely to impede marketers' efforts in any significant way. "The new targeting restriction actually aligns with Facebook's longer-term strategic playbook," explained one recent trade article, noting that "the type



of interest-level targeting to young people the company will now prohibit is slowly losing effectiveness when it comes to reaching all audiences.” The policy change, the article explains, “reflects Facebook’s own internal strategy shift away from advertiser self-guided targeting to campaigns that leave the decision-making up to the company’s almighty ad algorithm.” Rather than setting up campaigns that use targeting parameters to reach intended audiences, advertisers are being persuaded by Facebook “to let the company’s sophisticated algorithmic process for optimizing ad targeting do the work.”<sup>84</sup> Instagram also relies on the effectiveness of influencer marketing, a mainstay of the platform’s approach to advertising, to reach young users in ways that bypass more traditional online advertising tactics.

The heightened pressure and controversies have slowed Meta’s plans to encrypt Instagram. The company announced in a May 2021 blogpost that it was not planning to institute end-to-end encryption of its subsidiaries “any time soon.”<sup>85</sup> Acknowledging the “clear need to balance the privacy and security of people’s messages with maintaining a safe environment and providing data to law enforcement in response to potential real-world harms,” the company reported that it had not been able to reach consensus among the many stakeholders on this issue, “but experts encouraged more consultation to help strike an appropriate balance.”<sup>86</sup> However, in 2022, Meta announced it was conducting tests for instituting end-to-end encryption on Instagram Direct chats, setting the stage for encryption to “be available on all its messaging platforms.”<sup>87</sup>

## SNAPCHAT



Since its launch in 2011 by public company Snap, Inc., Snapchat has become extremely popular with teens. It was “ranked as teenagers’ favorite social media platform” in Piper Sandler’s “Taking Stock with Teens” survey for fall 2021, slipping into second place the following year as Tik Tok assumed more prominence.<sup>88</sup> The platform’s features and functionalities resonate very well with the personal and social lives of teenagers, enabling them to take photos on a mobile phone, personalize them with a toolbox of filters, and share them with friends, along with non-stop messaging, all of it largely under the radar of parents, caregivers, and teachers. Because it works with Wi-Fi, the app doesn’t use up a family’s data plan, and it is compatible with other devices, making it possible for teens who don’t have their own mobile phones to use the app. The ephemeral nature of Snapchat’s content and communications offers a unique digital cultural experience for teens. Any picture, video, or message that gets sent will be available only for a short time and then will disappear forever. As one industry insider characterized it, Snapchat is WhatsApp with selfies, where one can leave a message with

friends without leaving a trace. “It’s like beta testing yourself, pressure-testing your forming identity in real time.”<sup>89</sup>

These features have also made Snapchat particularly appealing to advertisers. The platform offers a robust digital marketing platform that leverages its virtual and augmented reality lenses and filter applications. The company promises marketers multiple ways to identify, track, and target individual users, which includes tapping into its unique Geofilters.<sup>90</sup> Snapchat is expanding its advertising services to offer new formats and techniques, including ecommerce sales. It has also been investing in neuromarketing-related research to understand how better to leverage its youthful audience on behalf of marketers.<sup>91</sup>

Snapchat’s terms of service prohibit accounts that “promote or distribute pornography,” and the company says it reports child exploitation to authorities. Other rules prohibit bullying, harassment, encouraging violence or dangerous behavior, impersonation, deceptive practices, false information, terrorism, hate speech and the like. Users can file a report with the “safety team using our in-app reporting feature or by completing [a] form.”<sup>92</sup> The platform also provides a Safety Center, where users can get “detailed instructions on managing your Snapchat experience.”<sup>93</sup> For content that “is likely to be seen by a larger audience, such as its ‘Discovery’ and ‘Spotlight’ features, Snapchat either curates or pre-moderates its content.” For “Spotlight,” according to a post by its newly hired “first” head of global platform safety, “all content is reviewed with automated tools, but then undergoes an extra layer

***However, according to independent assessments by journalists and NGOs, Snapchat's safety systems have not been that effective, with some of the flaws rooted in the very features that make the app so popular in the first place.***

of human moderation before it can be seen." The goal is to "control virality" to help lessen "the appeal to publicly post illegal or potentially harmful content."<sup>94</sup>

However, according to independent assessments by journalists and NGOs, Snapchat's safety systems have not been that effective, with some of the flaws rooted in the very features that make the app so popular in the first place. A review conducted by Canadian group C3P, for example, found that there were "no CSAM-specific reporting options for Snaps or stories." Users "can only report nudity or sexual content," with "no option to flag involvement of a child." Nor can a user "report an image sent within the chat function if the image has been uploaded from the sender's camera roll (versus being taken directly through the app)." Moreover, because the images in chat automatically disappear, reporting them is futile.<sup>95</sup> In 2019, the Sunday Times of London conducted an investigation of allegations that predators were "flocking" to Snapchat, which had become a "haven for child abuse." The investigation found thousands of cases where pedophiles were able to groom teens and get children to share indecent photos of themselves through the app.<sup>96</sup> As Adam Scott Wandt from John Jay College of Criminal Justice in New

York told the Sunday Times, it is the "self-destruct" nature of Snapchat's messages that "makes it difficult for the police to collect evidence," because "most Snapchat messages are unrecoverable after 24 hours, even by law enforcement with a warrant."<sup>97</sup>

To address such concerns, Snapchat has instituted a number of changes to its safety and reporting operations, setting up an in-house Law Enforcement Operations team "dedicated to reviewing and responding to law enforcement requests for data related to their investigations," and establishing a new "Head of Law Enforcement Outreach." As the company explains on its blog, although "content on Snapchat is ephemeral... we have long offered law enforcement agencies the ability to, consistent with applicable laws, preserve available account information and content for law enforcement in response to valid legal requests." In October 2021, the company held its "first-ever Snap Law Enforcement Summit," with more than 1,700 state, federal, and local officials participating.<sup>98</sup> According to its July 2021 transparency report, Snapchat's Trust and Safety Teams use the same automated tools as other social media companies to detect and report CSAM, including PhotoDNA and Google's

Child Sexual Abuse Imagery (CSAI) Match, reporting that during the first half of 2021, "5.43% of the total number of accounts enforced globally contained CSAM," and that "we proactively detected and actioned 70 percent of CSAM violations."<sup>99</sup>

Testifying before the Senate Commerce Committee's Subcommittee on Consumer Protection, Product Safety and Data Security, Snapchat's VP of Global Public Policy Jennifer Stout characterized Snapchat as "an antidote to social media," and described its parent corporation as a "camera company." Its "architecture," she explained, "was intentionally designed to empower people to express a full range of experiences and emotions" by "connecting people who were already friends in real life."<sup>100</sup> As a consequence, the company maintains, it is inherently safer than many other social media platforms. Its default features for teens include banning public profiles for under-18 users, employing age-gating tools to prevent those users from seeing age-regulated content and ads, and offering "consistent and easy-to-use controls like turning off location sharing by default and offering streamlined in-app reporting for users to report concerning content or behaviors to our Trust and Safety teams." Like other platforms with high numbers of teens, Snapchat is also developing new safety features to address the heightened concerns over young people and the online environment. For example, the company recently announced it was "rolling out a feature to limit the discoverability of minors in Quick Add (friend suggestions)," and developing "tools that will give parents more oversight

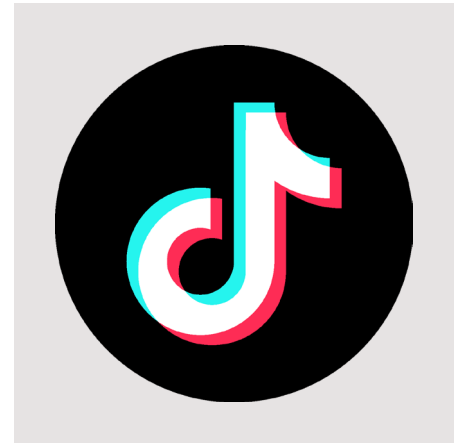
without sacrificing privacy,” which includes “plans to provide parents the ability to view their teen’s friends, manage their privacy and location settings, and see who they’re talking to.” Responding to concerns over social media and teens’ mental health, the platform also announced new educational initiatives, such as the “Here for You” program, which “provides support to users who may be experiencing mental health or emotional crises by providing tools and resources from experts.”<sup>101</sup>

However, there is continuing evidence that young people remain at risk on the platform. For example, in the summer of 2022, an analysis conducted by the Canadian Centre for Child Protection (C3P) found that teen boys were being subjected to “sextortion” tactics on both Snapchat and Instagram. “This is an ongoing problem that is getting worse,” noted Lianna McDonald, C3P’s Executive Director, “so it really does beg the question about what are these companies doing to keep children safe?”<sup>102</sup> Within a few days of the report’s release, Snapchat announced that it was launching a new “Family Center” application designed to “help parents get more insight into who their teens are friends with on Snapchat, and who they have been communicating with, without revealing any of the substance of those conversations.” The social media platform also announced plans to add additional features during the fall, including enhancing the “ability for teens to notify their parents when they report an account or a piece of content to us.”<sup>103</sup>

Snapchat has been working with IAS, DV and other companies since 2017 to expand its brand-safety operations. Currently, the company says it relies on both machine learning and human moderation to protect its advertisers, explaining that “We’ll also be moderating certain Snaps and Stories from popular Snapchatters in an effort to make sure that Snap ads and Promoted Stories don’t show up next to something that could be harmful to your brand. If a Snap or Story is flagged as unsafe for advertisers, then your ads will not show up in that Story, and that Story shouldn’t share the same screen as your Promoted Story on Discover.”<sup>104</sup> (Snap says it has “partnered with best-in-class safety experts like ConnectSafely and the UK Safer Internet Centre,” groups that are said to have helped build the company’s Safety Center).<sup>105</sup>

For Snapchat, the stakes of striking the correct balance between safety and monetization are especially high, as it faces growing competitive threats with its teen and Gen Z user base from TikTok, as well as from other giants such as Instagram and YouTube. To ensure a steady stream of ad dollars, it understands that it must make its safety operations work as effectively as possible, if it is to survive. But, as with other platforms catering to large numbers of youth, Snapchat’s safety features can sometimes be in conflict with or undermined by its commercial imperatives. A 2021 study conducted for 5 Rights in the UK examined Snapchat, among other companies with significant youth users, and identified a number of ways it can place youth at risk.<sup>106</sup>

## TIKTOK



TikTok has quickly captured the interest of children, adolescents, and young adults in 150 countries around the world. The mobile app enables users to create short video clips, customize them with a panoply of user-friendly special-effects tools, and then share them widely through the platform’s vast social network.<sup>107</sup> According to eMarketer, there were 2.3 million children under 11 years and 16.4 million 12-17 year-olds on TikTok’s U.S. platform in November 2021, and these numbers were continuing to grow in 2022.<sup>108</sup>

The key to TikTok’s success is how it has connected its music and video production and sharing applications with a highly sophisticated artificial intelligence system, designed to circumvent users’ conscious decision making, offer a constant stream of highly tailored content, and foster continuous interaction with the platform. With the initial launch of TikTok in 2016, China-based parent company ByteDance incorporated “deep learning algorithms” into the core of the app’s operations. Using computer vision technology to reveal insights based on images, objects, texts, and natural-language

processing, the app “learns” about an individual’s preferences, interests and online behaviors, so it can offer “high-quality and personalized” content and recommendations. “TikTok’s recommendation engines are distinct from algorithms used by Facebook and traditional networks, which rely heavily on suggestions from friends,” explained one business trade article. Modeled after ByteDance’s highly successful news app, TikTok “harvests insights based on what its users actually click on, read, and watch—right down to the type of music, faces, and voices in videos—learning as it goes. The distinction is subtle but important. Facebook was revolutionary because it tapped into our social grid of friends to serve relevant recommendations, but TikTok goes right to the source using AI to map out interests and desires we may not even be able to articulate to ourselves.”<sup>109</sup>

TikTok has worked assiduously to develop an arsenal of data-marketing technologies. The company deploys a range of targeting and brand-promotion applications that take full advantage of its unique AI and machine-learning system, enabling advertisers to tap into a vast store of user information, including not only age, gender, location, and interests, but also granular data sets based on constant tracking of behaviors and activities on the platform.<sup>110</sup> Marketers can target using a number of “creative ad formats,” such as “hashtag challenge” campaigns, “playable” ads, “smart videos,” and paid influencers, to name only a few.<sup>111</sup> Advertisers can also take advantage of the app’s “Automated Creative Optimization” tool, which “automatically” delivers “high-performing combinations” of marketing content.<sup>112</sup> Through internal

***The key to TikTok’s success is how it has connected its music and video production and sharing applications with a highly sophisticated artificial intelligence system, designed to circumvent users’ conscious decision making, offer a constant stream of highly tailored content, and foster continuous interaction with the platform.***

and external measurement services, marketers can receive detailed reports on the impact of their advertising techniques on individual users.<sup>113</sup> TikTok has continued to expand its digital marketing offerings in the U.S., working closely with a number of marketers that target youth.<sup>114</sup>

Since its initial launch, TikTok has attracted considerable controversy. For example, concerned about possible security breaches, U.S. government authorities launched a national security review of Chinese owner ByteDance when it acquired TikTok’s predecessor app, Musical.ly.<sup>115</sup> In 2019, a BBC investigation found hundreds of sexually explicit comments on videos that had been posted by children as young as nine years old. Even though TikTok claimed it had deleted the comments, it allowed the users who had posted the comments to remain on the platform, in violation of its own rules.<sup>116</sup> Also in 2019, The Intercept reported that TikTok had engaged in widespread discrimination against certain individual users whose representation was considered undesirable. The publication accused TikTok of instructing its moderators “to suppress posts created by users

deemed too ugly, poor, or disabled for the platform.” The Intercept also published internal TikTok moderation documents, which included instructions for barring such content and explained how the platform engaged in “algorithmic punishments for unattractive and impoverished users.”<sup>117</sup> A coalition of groups focused on youth privacy also pressed the FTC and the Department of Justice to investigate TikTok’s extensive data and targeting infrastructure designed to attract and interact with teens.<sup>118</sup>

The company issued a public statement in 2019 noting that these controversies had prompted it to institute a number of safety protocols: “Over the past several weeks, TikTok has been the focus of public statements by Members of Congress and others regarding alleged censorship of content, as well as questions about our data protection practices.... [W]e take these issues incredibly seriously as well. We are committed to transparency and accountability in how we support our TikTok users in the US and around the world.”<sup>119</sup>

In the last few years, TikTok has undertaken a series of ambitious efforts to allay public concerns over



its content, security, and safety problems, and to offer assurances to government authorities that there is no need for regulation. In March 2020, it established a U.S. content-moderation committee, the “Content Advisory Council,” to focus on “platform integrity, including policies against misinformation and election interference.”<sup>120</sup> It also set up three “transparency centers” in Los Angeles, Washington, D.C., and Dublin, in which “invited guests have the opportunity to see up close how we moderate and recommend content, secure our platform, and protect people’s privacy,” and can “review source code and learn how our application’s algorithm operates.” These centers have not yet opened publicly due to the coronavirus pandemic. However, in the meantime, the company offers “virtual tours for parts of the experience.”<sup>121</sup> In November 2020, the company announced that it was significantly expanding its content-moderation staff, hiring workers away from other social media networks and third-party firms such as Accenture and LCPL, “scaling the number of content moderators it has,” which now amount to 10,000 people worldwide.<sup>122</sup> TikTok also has a regional Trust & Safety hub in Singapore.<sup>123</sup> In 2022, the company announced a series of measures to provide “selected researchers” and members of the advisory council access to information on its content moderation and other safety-related processes.<sup>124</sup>

TikTok partners with Integral Ad Science (IAS), which regularly provides the popular platform with a “curated pool of videos to analyze” and categorize using the “GARM Brand Safety and Suitability categories.” This review process, performed prior

to the launch of an ad campaign for clients of IAS, is conducted “multiple times a week.” IAS identifies any TikTok content that should not receive advertising (the so-called GARM “floor” baseline), and also generates a “low, medium or high” risk rating for other material.<sup>125</sup> TikTok is also working with other brand-safety companies, including Zefr, which uses a “cognition AI machine learning engine... audio, text, and video frame-by-frame analysis” to assess ad placement reflecting “each of the 11 GARM categories.”<sup>126</sup>

In its public statements and on its website, the company assures regulators, the public and its users that it has established the most rigorous of standards and safety operations. It partners with a number of child-safety NGOs, including the Family Online Safety Institute, ConnectSafely, the Internet Watch Foundation, and Common Sense Networks, a for-profit affiliate of Common Sense Media.<sup>127</sup> It also recently created a “limited interactive experience” for U.S. children under the age of 13, enabling them to “engage with TikTok’s fun video features while limiting the information collected from them,” in order not to run afoul of the Children’s Online Privacy Protection Act (COPPA).<sup>128</sup>

TikTok has established a set of safeguards especially for its teen users, including “a neutral, industry-standard age-gate that requires people to fill in their complete birthdate to discourage people from simply clicking a pre-populated minimum age.”<sup>129</sup> Its safety features for teens include a default that designates all accounts for youth under 16 as private, so that “only people you’ve approved as followers can see

your content,” as well as a Family Pairing feature, “which lets guardians manage a teen’s account, including setting restrictions on screen time, search and direct messages.”<sup>130</sup> Also offered are a series of guides, including a Guardian’s Guide that “breaks down how the app works and outlines best practices for keeping users safe,” and a Bullying Prevention Guide, “developed in partnership with the Cyberbullying Research Center, Net Family News and other organizations,” offering “tips on identifying bullying behavior, ways to prevent harassment on the platform, and where victims can find help.”<sup>131</sup> The company has also instituted a set of new protocols for push notifications, based on a user’s age. On accounts for children between the ages of 13-15, push notifications will not be sent after 9 p.m. and for those between 16 and 17, the notification will be disabled after 10 p.m. In addition, users between the ages of 16 and 17 will have to actively switch their settings if they want to enable direct messaging, while younger teens will not be able to use direct messaging at all.<sup>132</sup> Trade publication eMarketer reported in November 2021 that “TikTok has released a report with Praesidio Safeguarding, an independent agency, tracking the popularity and impact that viral self-harm-related content and dangerous ‘challenges’ on the app have on teens and children”<sup>133</sup>

In its continuing efforts to burnish its own safety credentials, TikTok has formed alliances with other social media companies and child-protection organizations. For example, in 2020, TikTok proposed a “global coalition” to foster collaboration among tech companies, governments, and NGOs.” According to the company,

TikTok's interim head, Vanessa Pappas, sent a letter to the heads of nine social and content platforms proposing a Memorandum of Understanding (MOU) that would encourage companies to warn one another of such violent, graphic content on their own platforms. By working together and creating a hashbank for violent and graphic content, we could significantly reduce the chances of people encountering it and enduring the emotional harm that viewing such content can bring—no matter the app they use.<sup>134</sup>

Additionally, in October 2021, TikTok joined forces with INHOPE, a global network of 47 hotlines in 43 countries that provide the public with a way to anonymously report illegal content online with a focus on CSAM. According to the Netherlands-based NGO, “By establishing a cooperative partnership with INHOPE, TikTok joins companies like Google, Facebook, and Twitter, to directly support INHOPE’s work in Technology development, Network expansion, Capacity building, training, knowledge sharing, and Political/legislative engagement.”<sup>135</sup>

However, none of these strategies has been able to shield the company from regulatory oversight. In March 2022, a bipartisan group of attorneys general launched an investigation into TikTok’s potential harms to the mental and physical health of young people.<sup>136</sup> The UK’s data protection commission also announced it was investigating the social video platform for having failed “to protect children’s privacy.”<sup>137</sup>

## TWITCH



Originally launched in 2011, and acquired by Amazon in 2014, Twitch has become a highly successful sports and gaming platform.<sup>138</sup> Twitch is also home to a booming enterprise of amateur online gamers who see themselves as “entertainers,” “creating content for an audience who enjoys watching,” which is one of the latest and most consequential trends in the industry. These gamers—also known as “streamers”—are good at what they do, and they can attract tens of thousands of fans who like to see them play.<sup>139</sup> They broadcast their “feeds” on personal channels available to them on Twitch, where “fans watch and interact through chat windows and other platform features that allow them to have 1:1 communication with the personality they follow, as well as with

fellow fans.”<sup>140</sup> The Daily Beast reported that half of Twitch users said they were spending more than 20 hours a week on the platform.<sup>141</sup>

Twitch is at the center of a burgeoning new cultural phenomenon that has captured a large swath of the youth demographic. The site reports that 70 percent of its viewers are between 16 and 34 years of age.<sup>142</sup> According to the World Advertising Research Center (WARC), streaming on gaming platforms is now considered the “new prime time” for young audiences.<sup>143</sup> Gaming generates more revenue than TV, film or music, attracting viewers and players who are “highly engaged for a considerable length of time.” U.S. consumers spent \$51 billion for gaming software and services in 2021, while the total gaming industry was slightly more than \$180 billion worldwide (predicted to be \$200 billion in 2023).<sup>144</sup> Many teens and young adults report that they are building friendships through their online involvement as gamers or spectators.<sup>145</sup> “Contrary to the traditional stereotype that gaming is solitary and antisocial,” explained one industry study, “multiplayer games... are highly social environments that encourage collaboration, conversation and real-life connections.”<sup>146</sup>

**According to the World Advertising Research Center (WARC), streaming on gaming platforms is now considered the “new prime time” for young audiences.**

Twitch maintains the dominant lead in the video streaming gaming industry, with a reported average of 30 million daily viewers. According to industry tracking, the platform was responsible for hosting more than 91 percent of gaming streams.<sup>147</sup> As of February 2022, Twitch reported that more than 7 million unique streamers go live every month.<sup>148</sup> “Since the pandemic, Twitch’s userbase and market share has soared, with its hours watched increasing by 50 percent just between March and April,” reported one trade article. “It has long been the largest online streaming platform, but is increasingly attracting more than just gamers to stream.”<sup>149</sup>

Amazon integrated its Twitch gaming service into its overall advertising system in 2021, and now delivers an array of data- and content-driven marketing and advertising applications. Amazon is the third-ranking online ad platform (after Google and Facebook). The ecommerce and streaming giant provides marketers with a comprehensive and cutting-edge set of advertising opportunities, including for Twitch. Ads can “incorporate advertisements in live broadcasts across desktop, mobile, tablet, and connected TV devices,” which are “unskippable and always above the fold.” These advertisements cannot be stopped with ad blockers, according to Amazon. “Non-gaming” content on Twitch, such as its “Just Chatting” content, is growing. Influencer marketing is a key component of Twitch, which it encourages in order to promote the interests of its advertisers. Advertisers can leverage data and analytics from Amazon’s Marketing Cloud to trigger more effective campaigns. Underscoring how important video is overall for

***Twitch has been the site of a number of controversies involving sexual harassment and abuse, a problem that for years has plagued the online gaming culture, which has long been dominated by male players.***

marketers, Twitch is considered one of Amazon’s streaming video platforms, which also includes its “Freevee” TV service (formerly IMDb TV).<sup>150</sup>

Twitch has been the site of a number of controversies involving sexual harassment and abuse, a problem that for years has plagued the online gaming culture, which has long been dominated by male players. During the highly publicized “Gamergate” scandal, which began in 2014, female game developers, players, and critics were subject to extensive harassment, trolling, and even death threats by male gamers, forcing some of the women to seek protection.<sup>151</sup> In 2020, The Verge reported that dozens of people had recently reported sexual harassment, abuse, and assault on gaming platforms, mostly by males against females, with Twitch being singled out.<sup>152</sup> Most of the harassment came not from anonymous interlopers, but from some of the most popular male streamers with followers ranging from the thousands to the hundreds of thousands.<sup>153</sup> Many of those accused of this behavior were male Twitch “Partners,” streamers who have become successful enough that they are afforded numerous perks from the platform, “from a purple checkmark showing a streamer is verified to promotional opportunities at events like TwitchCon.”<sup>154</sup> Victims of this harassment accused Twitch of inaction,

arguing that “the company is giving these men a powerful position from which they can prey on fans or other streamers and that Twitch is failing to ban verified streamers despite credible reports of harassment or assault,” the article explained.<sup>155</sup> Adult streamers were also accused of predatory behavior toward children, in some instances taking advantage of Twitch’s own tools, such as its Whisper feature, which enables users to send private messages to one another across the entire platform.<sup>156</sup>

Although Twitch issued public statements promising to take the allegations seriously and look into the matter, those reporting the abuses remained concerned that the company was not doing enough.<sup>157</sup> Part of the problem stems from the nature of the Twitch platform itself, including both its technological affordances and its business model, which relies on “content creators” to amass large followings, which can, in turn, attract advertisers.<sup>158</sup> Given the considerable profits Twitch garners through these influential streamers, the company may be reluctant to punish or ban them from the platform. And because streaming and chat are both live, it is not possible to review what is broadcast in advance. Many of the abusive incidents are not occurring on the platform or even in public view; perpetrators often follow the women offline, including when they

go to TwitchCon or other streaming conventions, where they can be assaulted in person by the individuals who have been harassing them online.<sup>159</sup> The lack of effective platform policies, the spontaneous nature of the online streaming environment, and the company's unwillingness to penalize its lucrative star performers have all combined to create what one woman called a "Wild, Wild West," where victims of sexual harassment and abuse have little recourse and very few tools to protect themselves.<sup>160</sup>

Twitch's approach to online content moderation has made it challenging to address these kinds of problems. While moderating the streaming content through a centralized process, the company relies on its content providers to moderate the chat function within their streams, which is where the live conversations between gamers and their followers take place. According to one academic article, this system "allows for streamers to play an active role in governance and presents opportunities for more 'grassroots' moderation practices in comparison to centralized moderation."<sup>161</sup> The platform's guide suggests that content creators build their own "moderating team," by choosing people they know and trust to take on this role. "As a broadcaster," the guide explains, "choosing a good moderation team is one of the best things you can do for your channel," which is where the community-building, social interaction, and dialogue take place. "Moderators (also known as mods) ensure that the chat meets the behavior and content standards set by the broadcaster by removing offensive posts and spam that detracts from conversations. Mods can be easily identified in chat by the green sword icon that appears next to their name."<sup>162</sup>

***Twitch's community guidelines pointedly explain to its content creators (or "broadcasters") that they hold the responsibility for ensuring that content on their channels is moderated effectively, and that hateful conduct and harassment are swiftly addressed.***

Twitch's community guidelines pointedly explain to its content creators (or "broadcasters") that they hold the responsibility for ensuring that content on their channels is moderated effectively, and that hateful conduct and harassment are swiftly addressed. If they fail to handle these problems, they run the risk of being taken off the platform:

Creators are role models and leaders of the communities they create or foster around them. Creators should consider the consequences of their statements and actions of their audiences; sometimes unwanted focus or attention can encourage others to escalate their behavior into abuse.... You are expected to use channel moderators and the tools we provide, such as AutoMod, chat timeouts and bans, to mitigate hateful conduct and harassment in your channel. Not using the provided tools to moderate hateful conduct appearing on your channel will lead to a suspension. For harassment, broadcasters who do not make a good faith effort to specifically moderate incitement or organization of abuse in their channels through the tools provided will be suspended.<sup>163</sup>

With the sudden and dramatic increase in viewership during the pandemic, along with recent reports and bad press over sexual abuse and harassment, Twitch has had to take a number of public actions to deflect criticism and assure policymakers that the company can effectively police the platform. These include increasing its moderation staff, establishing an advisory council, strengthening its community standards, and expanding the use of automated machine-learning tools to flag abusive behavior. In May 2020, the company announced it was creating a new "Safety Advisory Council" to "inform and guide decisions." Among its duties are drafting and updating policies, "developing products and features to improve safety and moderation, promoting healthy streaming and work-life balance habits, protecting the interests of marginalized groups," and "identifying emerging trends that could impact the Twitch experience."<sup>164</sup> In November of that year, the company reported that it had doubled the size of its safety operations team and hired a new head of its Trust and Safety division, Angela Hession, who had worked for 20 years at Microsoft, where she spearheaded the establishment of Xbox's Project Artemis, described as a "grooming



detection' program to identify online predators attempting to lure children for sexual purposes."<sup>165</sup> In 2021, Twitch implemented a revised "Hateful Conduct and Harassment Policy," establishing specific guidelines for "harassing behaviors that are sexual in nature."<sup>166</sup> Its Safety team also recently posted a description of the ways it prevents children ages 12 and younger from being on the site, including the role of "automated solutions that use behavioral or language cues to identify users who are likely under 13 and proactively flag such accounts."<sup>167</sup>

Twitch's parent company, Amazon, has also established a number of brand-safety tools. "Pornographic or sexually explicit content, illegal content and extreme violence," among other categories, cannot be supported by ads, the ecommerce and gaming giant asserts. In 2021, Twitch platform observers reported that it had added a "brand safety score," described by some as an "experiment," aimed at grading "how brand friendly every streamer is based on things like chat behavior, ban history, manual ratings by Twitch staff, games played, age... and more."<sup>168</sup> In 2021, Amazon announced that advertisers could use the DoubleVerify system for a portion of its ad inventory. Amazon has not been as public, however, in disclosing how its brand-safety apparatus works on Twitch content.<sup>169</sup>

The 2020 Transparency Report (which it calls "the first of its kind for Twitch") explains the platform's "layered approach to safety," combining "the efforts of both Twitch (through tooling and staffing) and members of the community, working together." At its foundation are the content creators, who are required to learn

Twitch's Community Guidelines, "which seek to balance user expression with community safety, and set the expectations for behavior on Twitch." Creators are then expected to apply these "Twitch service wide standards" in the moderation of their own channels, and are free to "set an even higher bar if they choose, in their channel." Twitch provides content creators with "tools to set, communicate and enforce the minimum required standards of behavior in their channel."<sup>170</sup> The platform also makes tools available to viewers, enabling them to "customize the safety of their experience." These include "mature flags, chat filters, and blocking other users—that they can use to customize content and interactions they encounter across the service."<sup>171</sup> In addition, Twitch deploys "various technologies to proactively detect and remove certain kinds of harmful content before users ever encounter it. Finally, we empower users to report harmful or inappropriate behavior to Twitch. These reports are reviewed and acted on by a team of skilled and trained professionals who can apply service-wide enforcement actions."<sup>172</sup>

There is still skepticism within the child-protection community about Twitch's ability and willingness to offer effective safeguards for youth, women, and others who face risks on the platform. As the National Center on Child Sexual Exploitation noted, even with its strict policies prohibiting nudity, sexual exploitation, violence, harassment, and other harmful behavior and content, "Twitch has a serious problem with moderating and enforcing those policies in practice for the millions of viewers and streamers on their platform." Many of them

are quite young, and there have been instances of children as young as 11 revealing their ages directly on camera. Yet the platform has not put adequate age-verification or policing systems in place. "Despite the meteoric rise in users and the popularity of Twitch with the young demographic," the organization observed,

Twitch has yet to develop a set of lockable parental controls for minor accounts, leaving many parents frustrated and children unprotected. The current privacy and safety features are bare minimum at worst and confusing at best. Moderation is too often left to the streamers themselves to report bad actors and suspicious behavior. And when accounts do violate policies, like when one popular streamer accidentally featured child nudity, Twitch can choose to reinstate revenue-making users, choosing profits over policies.<sup>173</sup>

In 2022, Bloomberg reported on a detailed study that analyzed databases of Twitch accounts, identifying hundreds of "predatory accounts" set up by users primarily to "catalog, watch and manipulate children, including enticing them to perform everything from suggestive dances to explicit sexual acts." After conducting its own verification of the analysis, Bloomberg concluded that "Twitch's existing moderation tools have proved insufficient at preventing children from broadcasting and at stopping adults from finding and grooming them."<sup>174</sup>

Aware that there are growing concerns about the potential harmful impact Twitch can have on youth welfare, the company is investing in a "Community

***In 2022, Bloomberg reported on a detailed study that analyzed databases of Twitch accounts, identifying hundreds of “predatory accounts” set up by users primarily to “catalog, watch and manipulate children, including enticing them to perform everything from suggestive dances to explicit sexual acts.”***

Health Team” to “support the safety of millions of Twitch users.” It is hiring data scientists to help build “automated machine learning models that serve as the first line of defense against harmful behaviors,” including “quantifying the prevalence of harmful actions across the service.”<sup>175</sup>

While continuing to promote its youth safety protections to the public, Twitch touts the benefits of its platform to marketers seeking to reach and influence young people. “Advertising inside games offers all the same advantages as social media—targeting, measurement and increased brand awareness,” explained one trade publication, “but with greater creative freedom and a more immersive environment that garners focus and attention.”<sup>176</sup> For example, major brands for snacks, sugar-sweetened beverages, energy drinks, and fast food have eagerly embraced the platform, offering rewards to Twitch aficionados who engage with a “live streamer’s event,” and seamlessly integrating their brands into virtual gaming worlds.<sup>177</sup> Most of the marketing is integrated seamlessly into the game play, with streamers signing deals with brands that can sponsor their channels.<sup>178</sup>

Yet, it does not appear that Twitch has instituted any specific policies for advertising and marketing to teens. Its ad policies focus mainly on such standard issues as fairness, deception, and accuracy.

But while it offers no advertising safeguards to its users, Twitch includes a section on “Advertiser Safety” in its transparency report, underscoring the fact that the company feels the need to protect advertisers themselves from the kinds of content and behavior that also affect its users. “Advertising is an important part of Twitch,” the company declares, “and brands that advertise on Twitch want to know how we are making our users safer, and promoting a more positive and less harmful environment. As a condition of advertising with us, they want to ensure that their brand is not being associated with content or conduct that doesn’t align with their brand values.”<sup>179</sup>

Amazon’s efforts to develop new content-moderation policies and brand-safety tools for Twitch are directly related to a spectrum of regulatory and legal pressures it faces that have little to do with the operations of its popular gaming

platform. These include a growing anti-trust movement that seeks to break up the ecommerce giant, as well as a host of other political and legal challenges.<sup>180</sup> A coalition of advocacy groups, unions and small businesses are working together to force the company to change many of its practices, ranging from its dealings with independent and third-party retailers to how its labor practices affect workers and their efforts to form unions. Amazon clearly realizes that it will be even more politically and financially vulnerable if there are problems arising from its youth-oriented Twitch division.

## YOUTUBE



Initially launched as a dating site in 2005, YouTube quickly morphed into a more general platform where people could upload and share videos. Google purchased YouTube in 2006 and introduced advertising. It also created a YouTube Partner Program, which enabled content creators who uploaded videos to share in the revenues generated by ads.<sup>181</sup> With the ensuing explosion of video content, YouTube began attracting a growing number of children, despite the fact that its official terms of service barred anyone under

13 from the platform. YouTube is now widely considered the “number-one kids entertainment brand,” a position it has held since 2016. As one leading youth-marketing specialist noted, “the omnipresent platform plays a central role in shaping the evolving behavior of kids.”<sup>182</sup>

Over the years, YouTube has been criticized—by its own content creators and others—for its vague guidelines, inconsistent enforcement actions, and failure to address problems of hate speech, harassment, bullying, and sexual exploitation. “At times, the company seems to police videos primarily in response to public outcry, which makes its decisions inherently haphazard and potentially malleable,” explained a 2018 article in *Wired*. “It’s a gargantuan task to moderate a site to which hundreds of hours of video are uploaded every minute—but it’s even harder if you don’t have clear, consistent rules about what’s allowed.”<sup>183</sup> The company’s “advertiser-friendly guidelines” stipulate that any content that does not meet the guidelines would be “demonetized.”<sup>184</sup> But these and other measures have not kept the platform from becoming embroiled in repeated controversies and advertiser boycotts.<sup>185</sup>

A 2019 *Wired* study documented how predators were using the comment section of YouTube videos featuring children to guide other pedophiles to similar content. This scandal led several big brands to suspend their advertising on the platform, which in turn prompted YouTube to disable the comments function on many of the videos with children in them. But the recommendation system, which is at the heart of YouTube’s basic operations and financial success, continued to

push the sexual material out to a vast viewership on the platform.<sup>186</sup> The *New York Times* explained what the study’s researchers learned in tracking a video posted by a 10-year-old girl depicting herself and a friend playing in a backyard pool:

YouTube’s automated recommendation system—which drives most of the platform’s billions of views by suggesting what users should watch next—had begun showing the video to users who watched other videos of prepubescent, partially clothed children, a team of researchers has found. YouTube had curated the videos from across its archives, at times plucking out the otherwise innocuous home movies of unwitting families.... In many cases, its algorithm referred users to the videos after they watched sexually themed content. The result was a catalog of videos that experts say sexualizes children.<sup>187</sup>

Parent and child advocacy groups have called out YouTube for years over its failure to protect children, and its refusal even to admit that young children were on the platform. In an attempt to allay these concerns, YouTube launched a YouTube Kids App in 2015, with ad-supported child-appropriate content, “built from the ground up with little ones in mind.” According to the company’s official announcement, the app “makes it safer and easier for children to find videos on topics they want to explore.”<sup>188</sup> But despite the addition of this kids-only service, large numbers of young children continued to flock to YouTube’s main platform.<sup>189</sup> And YouTube Kids itself has become the subject of controversy, amid

reports of children being exposed to manipulative, bizarre and other types of inappropriate or harmful content on the purportedly child-friendly app.<sup>190</sup>

In 2018, a coalition of privacy, consumer-protection, and child-advocacy groups filed a complaint with the FTC, charging that Google had been disingenuously claiming that its YouTube service was intended only for those aged 13 and older, even though it was widely known that the platform had become the number-one online destination for younger children. In September 2019, the FTC and the New York attorney general reached a landmark settlement with Google that included a \$170 million fine, along with an agreement that Google would make a number of changes to YouTube’s business practices, affecting both its U.S. and global operations. YouTube would no longer allow personalized, “behavioral” marketing on programming that targets children, and safeguards currently in place on its YouTube Kids app would apply to all child-directed content on its main YouTube platform.<sup>191</sup> The company also promised to substantially curtail the data it collects from children watching YouTube videos, and to take steps to drive kids from the main YouTube site to YouTube Kids, where parental consent is required. In addition to proactively revising the algorithm used to make recommendations to children, the company announced it would no longer permit comments and notifications on child-directed content.<sup>192</sup> It also instituted new policies limiting recommendations of videos that depict “minors in risky situations,” banning live-streamed broadcasts by children “unless they are clearly accompanied by an adult,” and adding AI systems to find and remove

inappropriate or illegal content.<sup>193</sup> In 2021, YouTube announced that it was setting the default privacy setting for ages 13-17 to private video posts.<sup>194</sup>

YouTube's content-moderation system appears to be structured along lines similar to that of other major social media companies. In addition to employing its own in-house staff, Google contracts with specialized third-party companies, including Accenture, which operates Google's largest content-moderation site in Austin, Texas.<sup>195</sup> According to the Washington Post, "YouTube's content policies have tended to attract less media attention and scrutiny than those of Facebook or Twitter," even though the video-sharing platform has more than 2 billion users worldwide and is "the internet's dominant hub for user-created videos of all kinds." Though YouTube played "a major role in misinformation campaigns, including the effort to discredit the results of the 2020 U.S. presidential election, and its recommendation algorithms have been implicated in leading some users down a path of radicalization," observed the Post, "...it has repeatedly ducked the brunt of backlashes that Facebook and Twitter have absorbed head-on." As a consequence, the article explained, "researchers say that YouTube remains one of the hardest social media platforms to study, because video is so difficult to analyze in bulk and the platform doesn't provide many tools to do so."<sup>196</sup>

The company's transparency reports offer some insights into how its system works, but also leave a number of questions unanswered. According to the report from April-June 2021, the company relies on "a combination of people and technology to flag inappropriate content." As it explains,

***The company's transparency reports offer some insights into how its system works, but also leave a number of questions unanswered.***

"Flags can come from our automated flagging systems, from members of the Trusted Flagger program (NGOs, government agencies, and individuals) or from users in the broader YouTube community." The company recently introduced a new measurement "data point" to track the effectiveness of its content-moderation efforts. Called the "violative view rate," it measures the percentage of "views" of videos that violate YouTube's rules. As Vox News reported, "for every 10,000 views on its social network—or at least during the last quarter of 2020—about 16 to 18 of those views are of videos that violate YouTube's rules, which currently forbid everything from hate speech to medical misinformation about Covid-19 to spam."<sup>197</sup> YouTube touted a 70 percent decrease in the violative view rate since 2017 as a major success based on improvements the company had made in its content-moderation systems, particularly through the use of AI. But as the Vox article pointed out, "YouTube's own reviewers—not independent auditors—decide what counts as a violation of YouTube's guidelines."<sup>198</sup>

Google/YouTube has a "four-pronged approach" to brand safety, including removing content that violates its community guidelines; reducing the spread of potentially harmful content; elevating "authoritative voices"; and "rewarding trusted creators with the privilege of monetizing" their content.<sup>199</sup> Google operates its brand-safety system,

including for YouTube, through its "Display and Video 360" ad operations, which enable its advertisers to use the platform's own content-ratings system to decide whether to target an ad. Content labels mirror those in the film and television industries, ranging from "DL-G," which is "suitable for general audiences," to "DL-MA," the label for "mature audiences." Under its "Google Ad Manager" service, which has its own set of content policies, the company provides information on ways marketers and publishers can engage in practices that protect their brands, including child safety ("child sexual abuse and exploitation").<sup>200</sup> While Google's policies restrict content with sexual themes, there are still a number of exceptions where such content can run as ads.<sup>201</sup> Google's Transparency Center includes a separate area for "YouTube Creators," providing information on its policies and guidelines "designed to keep our community of Creators, viewers, and advertisers protected." If YouTube creators want to make money from ads, they must adhere to its "Guidelines and Monetization Policies," which require agreeing to a set of "Advertiser-friendly content guidelines" that prohibit (with some caveats) material involving adult and violent content from being supported by advertising. The company also has a separate set of policies governing what kinds of products can be advertised, as well as rules to protect the privacy of minors.<sup>202</sup>



## QUESTIONS AND STRATEGIC OPPORTUNITIES

The brief profiles presented above provide only a partial picture of how these five platforms are responding to the pressures they face. In the largely unregulated tech industry, social media companies have developed a multiplicity of strategies and tactics, mobilizing the full spectrum of content-moderation systems, corporate responsibility operations, brand-safety tools, and other institutional mechanisms to demonstrate that they are capable of policing themselves. While adding personnel to their trust and safety and content-moderation operations, social media companies are also relying more and more on automated systems, deploying AI and machine learning. From industry accounts of these changes, the systems that are now in place appear to be much more expansive and robust than they were only a year or two ago, capable of tracking and taking down CSAM and other illegal content, but also deployed to flag and remove a growing spectrum of controversial and problematic content, such as hate speech and vaccine disinformation. This further mechanization appears to have created a highly automated content-moderation system, which, when married to the technological capability of marketing and brand-safety operations, may have altered substantially the nature of social media, creating an online digital ecosystem that can engage in ongoing surveillance, instantaneous intervention, and highly granular content discrimination.

AI and machine learning are also at the heart of new strategies for addressing teen-safety issues. In one of the more interesting new developments, an increasing number of social media platforms are building into the navigation experience some new guardrails that structure and constrain teen interactions online. These kinds of defaults are not new; platforms such as Facebook, for example, instituted them in limited ways as part of their cyberbullying policies for teens.<sup>203</sup> However, with the recent pressures on the entire social media industry over how its operations are negatively affecting the mental health of teens, these automated tools have been adopted by more and more companies. As a consequence, they are becoming an increasingly pervasive presence in the social media ecosystem.

This is a remarkable development for an industry that has, for decades, tried to downplay concerns over teen vulnerabilities in the online environment, and which has fought legislative efforts to extend privacy protections to young people over the age of 12. The new teen-safety features are a response not only to rising controversies over social media's harmful impacts on teen users, but also to several recent U.S. and international policy developments that are affecting the operation of global tech companies. U.S. lawmakers have been pushing to extend COPPA's protections to teens, spurred by leading U.S. children's advocacy groups. The EU's General Data Protection Regulation (GDPR), adopted in 2018, which identifies a person under age 18 as a minor and

***While these new default teen-safety tools are a significant improvement, they also raise some important questions and concerns. For example, they appear to draw from the same arsenal of AI and other granular forms of surveillance, data analysis, and microtargeting that the industry has perfected for targeting teen users with marketing and advertising.***

thus deserving of special protections, has required online companies to tread more cautiously when it comes to dealing with teen-privacy issues. Criticism in China, India and other countries about the impacts on young people from online media has also sent a strong signal that government officials are concerned.<sup>204</sup> Finally, the UK's implementation of its Age Appropriate Design Code to protect young people online has had a rippling impact around the world, requiring leading companies to review how their products and services may affect young people.

While these new default teen-safety tools are a significant improvement, they also raise some important questions and concerns. For example, they appear to draw from the same arsenal of AI and other granular forms of surveillance, data analysis, and microtargeting that the industry has perfected for targeting teen users with marketing and advertising. To what extent are these two functions intertwined? Does the use of these safety tools expand the collection and use of data on teens? There are also questions about equity. Are all

teens affected by these tools equally, or might there be disparate impacts depending on gender, race, language ability, socioeconomic status, or other factors? For example, a number of the safety defaults are based on the assumption that parents and caregivers are in the best position to protect their own children from child sexual abuse material, predation and other harms. Yet this is not the case for all young people. There is research evidence that a significant percentage of child sexual abuse images is produced by family members or close, trusted adults.<sup>205</sup> Will the new safety tools protect these children or do them further harm? Children in low-income families may be at even greater risk. All of these systems need to be assessed independently to determine the extent and nature of their impact on youth.

It remains unclear whether social media platforms are conducting effective independent assessments of their various safety policies, tools, and default operations. Facebook has employed outside experts to conduct evaluations of its content-moderation systems and, along with several other social media platforms, has set up

external entities to help guide and adjudicate its content-moderation decisions. But most platforms do not provide any documentation of how or even whether they assess these safety systems. They may be conducting this research internally, but have not made the evaluations public yet. This may be worth investigating further.

Although platforms, marketers and digital media trade groups continually claim they are concerned about young people, the industry's efforts to establish safeguards are, at best, fragmented and conflicted. In most cases, companies have been careful not to allow any new safety systems to fundamentally interfere with their advertising practices or undermine their business models. Developing strong, monetizable, connections to children and teens remains an essential goal for platforms, whose ongoing profitability relies in no small part on the sales of games, toys, fast food, and entertainment to young users. Platforms and other online content companies are making major investments to expand the ways they can attract more young people, maximize their engagement with digital services, and generate ad and other revenues. The tech industry is also pouring hundreds of millions of dollars into building up its lobbying apparatus and contributing money to key politicians who can help them evade regulation.<sup>206</sup>

Even with the industry's recent moves toward greater transparency, most of the operations inside these companies remain hidden from public view, leaving many questions about how the various safety protocols and teen-friendly policies actually function. For example, how well and in what ways

do brand safety, content moderation, and by-design navigation tools work hand in hand? How much control do advertisers have over content? What kinds of responsibilities are being assumed by other participants in the online social media ecosystem? A key question is whether the platform and online industry's commitments to making the internet safer for advertisers is a more comprehensive and economically supported effort than what the companies do with content moderation and other efforts related to the safety of youth. Does brand safety overshadow child-safety efforts, given that the former's role is to enable successful ongoing revenue generation from advertisers? Are current brand-safety systems effectively preventing the monetization of CSAM-related sites?

Brand-safety systems deserve greater scrutiny by advocates and policymakers. Advertisers are a powerful force in the entire digital media ecosystem, yet their increasingly influential role in content decisions has remained largely absent from the public debates over children and the internet. The industry has established an expanding set of institutions, policies and technologies that are shaping the operations of the leading platforms and publishers, continually fine-tuning these systems to minimize risks while maximizing monetization. Can the affordances of the brand-safety complex, including ad targeting and analysis, be used to assess CSAM

***A key question is whether the platform and online industry's commitments to making the internet safer for advertisers is a more comprehensive and economically supported effort than what the companies do with content moderation and other efforts related to the safety of youth.***

and other harmful content more effectively? How well do the current mechanisms work now? The child-safety field needs to have an assessment of the investment and technology-deployment decisions platforms are making in AI and other content-recognition technologies. How are they being implemented in the emerging metaverse? How well do child-safety safeguards function within the broader brand-safety system, which is required to cover an increasingly diverse set of problematic content, including disinformation, hate speech, and other issues?

The brand-safety field is still in a state of flux. For the most part, the various sectors that comprise the advertising and technology industries are working together to advance their collective interests. However, there are a number of fault lines and conflicts that could create opportunities for child advocates to press for better protections on behalf of young people. The industry is in the midst hammering out a

number of "access rights and media governance" issues, with advertisers pressuring platforms to provide greater accountability for leading marketers. There is debate within the industry about how it should respond to rising public concerns over privacy, protecting children and countering hate speech and racism online. There is also increasing recognition and support for enactment of data-protection regulations, building on the EU's General Data Protection Regulation. Industry leaders are also responding to the growing chorus of children's advocates, former digital industry executives and policymakers decrying the addictive nature of digital media, including the manipulative role played by algorithms, which operate in a "black box" with their own hidden agendas and goals. All of this fluidity creates an opportunity for child advocates to engage in dialogue with key industry groups such as WFA and GARM, and to promote a stronger agenda on behalf of youth safety, privacy and health.

## CURRENT U.S. POLICY PROPOSALS ON CHILDREN AND TECHNOLOGY

The most effective way to ensure greater accountability and more meaningful transparency by the social media industry is through stronger public policies. As part of his March 2022 State of the Union address, President Joe Biden issued a strong indictment of the technology industry over its treatment of young people.<sup>207</sup> His language echoed many of the statements that congressional leaders had been making in the wake of the explosive revelations over Facebook’s internal research on Instagram and teens. Recognizing whistleblower Frances Haugen, who was an invited guest in the audience, the president accused “the large social media platforms” of “conducting a national experiment on our children and using their data to keep them clicking.” The industry’s practices are accentuating a “youth mental health crisis,” he declared, citing “mounting evidence that social media is harmful to many kids’ and teens’ mental health, well-being, and development.” To reverse this trend, Biden called for Congress to “strengthen privacy protections, ban targeted advertising to children, and demand technology companies stop collecting personal data on children.”<sup>208</sup> In a background document distributed prior to the speech, the White House presented a technology regulatory framework that would require platforms and other digital services “to prioritize and ensure the health, safety and well-being of children and young people above profit and revenue in the design of their products and services.” It also called for policies to “stop discriminatory algorithmic decision-making that limits opportunities for young Americans,” along with greater federal financial support to conduct “research on social media’s mental harms,” highlighting that “children under 18 are disproportionately vulnerable to the dangerous and harmful content they might encounter online.”<sup>209</sup>

The White House’s background document presented these proposed protections for youth as part of a broad strategy for addressing the “national mental health crisis,” listing more than a dozen policy goals that extend far beyond the tech industry, encompassing changes in treatment protocols, access to health systems, mental health parity in insurance, and more. Framing youth and technology issues as a mental health problem reflects a clear recognition of the power of today’s pervasive digital media environment and its influence on young people’s psycho-social, emotional, and physical development. The White House also knows that this topic has garnered wide bi-partisan support, part of what the Biden administration calls its “unity agenda,” consisting of “policy where there has historically been support from both Republicans and Democrats.”<sup>210</sup>

Whether this push from the White House ultimately results in passage of technology regulation remains to be seen. But it is clear that we are witnessing an unprecedented political environment, in which policymakers on both sides of the aisle are calling for government safeguards to protect children. At this point, there is strong interest in Congress for policies that would address three major areas related to the tech industry’s relationship to young people: updating the Children’s Online Privacy Protection Act (COPPA); instituting new design requirements to govern how young people interact with social media and other digital platforms; and strengthening safeguards against CSAM and other sexual abuse



practices, which includes reforming the immunity protections that tech companies now have under Section 230 of the Telecommunications Act. While versions of these bills have been introduced in prior Congressional sessions, some elements are new and others have been retooled and updated in light of recent controversies in the tech industry. We briefly review each of these areas below:

## PRIVACY

Children's privacy-related bills have been introduced in both the House and Senate, informally labeled "COPPA 2.0," which would update the landmark children's privacy law, first passed in 1998. In the Senate, the "Children and Teens' Online Privacy Protection Act," co-sponsored by Senators Ed Markey (D-Mass.) and Bill Cassidy (R-La.), would prohibit targeted advertising to children 12 and under, and would require an affirmative "opt-in" for ads targeted to youth aged 13-16. It would also establish a "Youth Privacy and Marketing" division at the FTC, and provide that agency with additional tools for enforcing the law. In the House, the leading, and stronger, measure to advance privacy protections for youth is called the PRIVCY Act ("Protecting the Information of our Vulnerable Children and Youth Act). It would apply to sites "likely to be accessed by children and teens" and would ban surveillance advertising to youth under 18. The House bill, borrowing from the UK's Age Appropriate Design Code concepts, also requires platforms to make the "best interest of children" their primary consideration when designing their services, among other provisions.<sup>211</sup>

In July 2022, the House Energy and

Commerce Committee voted to advance the American Data Privacy and Protection Act (ADPPA), H.R. 8152, to the full House of Representatives. This bipartisan bill would create a comprehensive consumer privacy framework with additional significant protections for individuals under 17. The bill would prohibit the collection, use, and transfer of covered data beyond what is reasonably necessary and proportionate to provide a service requested by the individual, unless the collection, use, or disclosure would fall under one of seventeen permissible purposes. Additionally, it would create special protections for sensitive data, which includes data of individuals under 17. Other safeguards include significant restrictions placed on data brokers and a prohibition to use covered data in a way that discriminates on the basis of protected characteristics (such as race or sex). While these safeguards would apply to all consumers, the bill would implement additional safeguards for minors. The bill would include a prohibition on targeted advertising, and it would establish a Youth Privacy and Marketing Division at the FTC. These additional protections would apply only when the covered entity knows the individual in question is under age 17, though certain social media companies or large data holders would be deemed to "know" an individual's age in more circumstances. Overall, if ADPPA were to pass, it would provide historic safeguards for minors across a range of online risks and would constitute significant improvements over the past.

## SAFETY, DESIGN AND RESEARCH

Also paralleling the UK's Design Code, the bi-partisan Kids Online Safety Act of 2022 (KOSA), co-sponsored by Senators Richard Blumenthal (D-Conn.) and Marsha Blackburn (R-Tenn.), is similarly aimed at ensuring that social media platforms "put the interests of children first." Applying to children under 17, KOSA is meant as a regulation of design features and is intended to work in conjunction with legislation, such as COPPA 2.0, which is focused on privacy safeguards. Platforms would be required to "enable the strongest settings by default" and provide "tools" to enable young people to "disable product features and opt out of algorithmic recommendations." The legislation would make it incumbent upon social media companies to "prevent and mitigate harms to minors," including protecting them from content related to "self-harm, suicide, eating disorders, substance abuse, and sexual exploitation." It would require these platforms to provide minors with options to protect their information and disable addictive product features. It would require tech companies to provide parental controls, including a mechanisms so that that would enable parents to learn about potential risks and easily report them to the platform; teens would be provided notice if parental controls are in effect. Platforms would be required to provide "access to critical datasets" that scholars and nonprofit groups could use to conduct research on "harms to the safety and well-being of minors."<sup>212</sup> The "Kids Internet Design and Safety Act" (known as "The KIDS Act"), introduced by Senator Markey

in the Senate and by Representative Castor in the House, would mandate that online platforms design their services to “prevent the amplification of harmful content” and “stop manipulative” features, including those used to target youth with advertising and marketing.<sup>213</sup> The “Children and Media Research Advancement Act” (CAMRA) would direct the National Institutes of Health (NIH) to conduct a five-year study to assess how technology impacts the healthy development of youth.<sup>214</sup>

## STATE-LEVEL DESIGN LEGISLATION

As Congress continues to debate these various legislative vehicles, California lawmakers have just passed their own “Age-Appropriate Design Code Act,” which will take effect in 2024. This landmark legislation, the first in the U.S. based on the UK code, covers children under 18 and requires companies to “consider the best interests of the children” as a condition for developing online services and products offered in that state. It requires online services “likely to be accessed by children” to assess the site’s risks for minors, limit using personal information from minors, avoid collecting geolocation data unless “strictly necessary,” implement privacy-by-default settings, and restrict the use of profiling. It also restricts the use of “dark patterns,” a general term for manipulative design features intended to influence online behavior.<sup>215</sup>

## REFORM OF SECTION 230

Several bi-partisan bills have been introduced that would amend Section 230 of the Telecommunications Act. For example, the EARN IT Act (“Eliminating Abusive and Rampant Neglect of Interactive Technologies Act”) would trigger the loss of Section 230 “blanket immunity” protections if online platforms failed to take proactive action to prevent posting of CSAM to their sites.<sup>216</sup> It would create a national commission composed of law enforcement, experts and abuse survivors to develop “best practices to address child abuse online.” The SAFE Tech Act would amend Section 230 by increasing the liability of platforms on issues such as cyberstalking, targeted harassment, wrongful death, and when paid content is involved.<sup>217</sup> The “Justice Against Malicious Algorithms Act” would amend Section 230’s immunity protections if “an online platform knowingly or recklessly uses an algorithm or other technology to recommend content that materially contributes to physical or severe emotional injury. The “Invest in Child Safety Act” would provide mandatory funding so that the Department of Justice, FBI and NCMEC would have greater resources to address online sexual abuse.<sup>218</sup>

While many of the bills for protecting children are being supported by a spectrum of liberal and conservative child-protection and children’s privacy organizations, proposals to amend Section 230 have, predictably, generated strong opposition from civil

liberties groups, LGBTQ advocates, and others who oppose any changes to that law. As the Washington Post observed, the EARN IT Act has “reignited a battle over the future of Internet regulation and online speech.” Opponents of the bill are warning that it could undermine privacy by disabling encryption and other technologies designed to make communications more secure. The Electronic Frontier Foundation issued a warning that the proposed law “empowers every U.S. state or territory to create sweeping new Internet regulations.... [They] will be allowed to pass whatever type of law they want to hold private companies liable, as long as they somehow relate their new rules to online child abuse.”<sup>219</sup>

## BUILDING A “DIGITAL ENABLING ENVIRONMENT” FOR YOUNG PEOPLE

The controversy over the EARN IT Act echoes similar, longstanding public debates over young people and the Internet, where the goals of ensuring the wellbeing and safety of children are often pitted against the goals of preserving privacy and guaranteeing free speech. These fault lines were also in evidence in the wake of Apple’s 2021 announcement of new children’s safety protocols, which elicited a powerful outcry from civil liberties and privacy groups, and ultimately caused Apple to retract its plans.<sup>220</sup> Moving forward with any coordinated agenda on behalf of children will require advocates to address these deep-seated fissures within civil society and to forge consensus on how we, as a society, can create an internet that treats all people, including children and teens, with fairness and dignity.

For years, tech policies in the U.S. have followed a narrow, piecemeal approach to addressing children’s needs in the online environment, providing safeguards for only the youngest children, and failing to take into account the holistic nature of young peoples’ engagement with the digital media environment. There is a need for more comprehensive and integrated policies that combine privacy protections and advertising and commercial surveillance safeguards with child-safety and CSAM protection. This will require a strategic effort that brings together the diverse constituencies working on behalf of youth in the online media.

Traditionally, children’s privacy organizations have operated within their own networks, and have not interacted with the community of activists focused on CSAM and related child-protection issues. Children’s privacy advocates have also worked within broader coalitions representing a range of privacy, civil liberties and other tech policy concerns, whose goals are often at odds with the those of child-protection advocates. But while the divides between these communities remain very wide, regulatory proposals on behalf of children continue to generate bi-partisan support in Congress and among the general public. The successful passage of

California’s design code law, and the growing support across a spectrum of communities for similar legislation at the federal level, could help lay the groundwork for a broader policy agenda. More could be done to build bridges between these divided advocacy communities in order to forge common ground on overarching values, and to find workable solutions beyond the entrenched polarization that currently exists. Human rights, fairness, justice, equity, and healthy child development must be moved to the center of the policy debate.<sup>221</sup> Because the impacts of digital technologies on children are so widespread, efforts should also be made to broaden the coalition of organizations that have traditionally fought for children’s interests in the digital media, to include groups representing the environment, civil rights, health, education, and other key stakeholder communities.

U.S. advocates should also leverage policy developments in Europe to hold the technology industry more accountable to children. The UK’s Design Code is already influencing policy development and corporate practices in the United States. U.S. advocates will have further opportunities to collaborate with European allies in order to develop stronger and more comprehensive strategies. For example, the recently adopted provisions of the EU’s “Digital Services Act” (DSA) stipulate that platforms “accessible to minors” will have to assess how their services interact with young people and “take specific measures to protect them.”

The rules cover a range of regulations concerning young people, including new procedures for takedown of CSAM and other illegal content, and prohibitions against targeted data-driven surveillance advertising to minors.<sup>222</sup>

While European regulations do not apply to tech companies’ U.S. operations, the combined pressure from stronger, comprehensive rules in the EU, and bi-partisan support in the U.S. Congress for new regulations—some of which are modeled on European policies—have continued to increase the pressure on the global tech industry. And all of this is happening at a time when the youth online marketplace is rapidly expanding into a booming “metaverse.” These developments have created a critical window of opportunity for advocates to engage in dialogue with industry leaders, while also supporting new policy and regulatory initiatives.<sup>223</sup> In addition to social media platforms, advertisers play a critical role in how young people are treated in the online environment. Advocates should meet with international leaders of the marketing industry, including members of the World Federation of Advertisers, to urge them to commit greater resources to supporting the interests of children and adolescents.

Drawing from the research for this report, we want to highlight several key issues that we believe advocates should consider as they develop a U.S. tech policy agenda for young people.

1) First is the need to establish industry-wide safeguards for tech companies engaging with children. Under the current self-regulatory system, each tech platform has been

able to develop its own internal corporate policies and practices, which are often designed to deflect criticism, preempt regulation, and safeguard prevailing business operations. These various safety “tools,” ad policies, content-moderation practices, and even transparency systems are too often difficult to understand, and are constantly undergoing changes. In their place, we should insist on uniform industry-wide standards that will ensure that children are afforded the same level of safeguards across the digital media ecosystem. These could be tailored to the features and affordances of individual platforms, but they should adhere to some measurable criteria for effectiveness. In addition, global technology companies should offer all children, no matter where they live, a similar set of protections. The UK Design Code provides a framework for the development of such standards, but policies would need to be specific enough to ensure uniformity across the industry.

2) These safeguards will need to be evaluated by independent entities on a regular basis, covering the full range of safety policies, ad practices, design codes, and privacy protections. The results should be made public and companies should be required to make any necessary changes to improve the effectiveness of their practices.

3) Regulatory agencies, advocates, and scholars will need to conduct ongoing analyses of the commercial practices and online operations targeting children and youth. As the digital marketplace continues to grow and evolve, technological and commercial innovations are constantly changing the way that data-

driven and engagement models are influencing young people in the digital environment. Regulations and industry safeguards will need to be reviewed in the context of these changes and altered accordingly to stay abreast of fast-moving developments.<sup>224</sup>

Finally, we suggest that stakeholders should work together to identify the kinds of outcomes we wish our children to experience from the online media culture. One approach would be to draw from the field of positive youth development, which calls for an “enabling environment” for young people, one that “supports their assets, agency, access to services, and opportunities, and strengthens their ability to avoid risks and to stay safe, secure, and be protected and live without fear of violence or retribution,” and that “encourages and recognizes youth, while promoting their social and emotional competence to thrive.”<sup>225</sup> Though this concept was not devised specifically for digital technologies, its principles and goals could become key pillars in a framework for a “digital enabling environment” for young people. The framework should also be rooted in the recommendations adopted by the United Nation’s Convention on the Rights of the Child (UNCRC), declaring that the rights of “every child” around the world must be “respected, protected and fulfilled in the digital environment.”<sup>226</sup> In addition to addressing the needs of the individual child, we need to take into account the unique experiences of the United States as a nation, ensuring that digital technologies produce equitable outcomes and enable equal opportunity for specific groups of young people, including those from marginalized socio-economic and racial/ethnic backgrounds.



# REFERENCES

- 1 Georgia Wells and Jeff Horwitz, "Facebook's Effort to Attract Pre-teens Goes Beyond Instagram," *Wall Street Journal*, 28 Sept. 2021, <https://www.wsj.com/articles/facebook-instagram-kids-tweens-attract-11632849667> (subscription required).
- 2 Vanessa Romo, "4 Takeaways from Senators' Grilling of Instagram's CEO about Kids and Safety," NPR, 8 Dec. 2021, <https://www.npr.org/2021/12/08/1062576576/instagrams-ceo-adam-mosseri-senators-brush-aside-his-promises-to-self-poli>.
- 3 Makena Kelly, "Congress Will Investigate Claims that Instagram Harms Teens," *The Verge*, 15 Sept. 2021, <https://www.theverge.com/2021/9/15/22675802/facebook-instagram-teens-harm-blumenthal-blackburn-markey-kids>.
- 4 Meta, "Connection is Evolving and So Are We," [https://about.facebook.com/meta?utm\\_source=Google&utm\\_medium=paid-search&utm\\_campaign=metaverse&utm\\_content=post-launch](https://about.facebook.com/meta?utm_source=Google&utm_medium=paid-search&utm_campaign=metaverse&utm_content=post-launch). As one online publication recently explained it, the emerging metaverse is "a more permanent virtual reality that bleeds into our physical lives." Michelle Santiago Cortés, "What is the Metaverse? A (Kind Of) Simple Explainer," *Mashable*, 9 Nov. 2021, <https://mashable.com/article/what-is-the-metaverse-explainer>.
- 5 Kathryn C. Montgomery, *Generation Digital: Politics, Commerce, and Childhood in the Age of the Internet* (Cambridge, MA: MIT Press, 2007).
- 6 Victoria Petrock, "US Generation Z Shopping Behaviors," *eMarketer*, 15 Nov. 2021, <https://www.insiderintelligence.com/content/us-generation-z-shopping-behaviors>.
- 7 Charles Gabriel, "What Marketers Need to Know about Generation Alpha," *WARC*, September 2021. <https://www.warc.com/content/article/bestprac/what-marketers-need-to-know-about-generation-alpha/engb/138894>.
- 8 Canadian Centre for Child Protection, "Reviewing Child Sexual Abuse Material Reporting Functions on Popular Platforms: Executive Summary," 2020, [https://www.protectchildren.ca/pdfs/C3P\\_ReviewingCSAM-MaterialReporting\\_Summary\\_en.pdf](https://www.protectchildren.ca/pdfs/C3P_ReviewingCSAM-MaterialReporting_Summary_en.pdf).
- 9 Michael Passoff, "Facebook At Center Of Storm Over Child Sexual Exploitation Online," *Proxy Preview*, 17 Mar. 2020, <https://www.proxy-preview.org/all-contributor-articles/2020/3/16/facebook-at-center-of-storm-over-child-sexual-exploitation-online>.
- 10 Tech companies appear to be skirting laws like the Children's Online Privacy Protection Act (COPPA), which were designed to protect children under 13. For example, a 2021 report from the child protection organization Thorn found that large numbers of young children were engaging with major social media platforms. Many of these children reported having sexual interactions, not only with peers, but also with individuals they believed to be adults. Casey Newton, "The Child Safety Problem on Platforms is Worse than We Knew," *Platformer*, 12 May 2021, <https://www.platformer.news/p/the-child-safety-problem-on-platforms>. See also, Thorn, "Responding to Online Threats: Minors' Perspectives on Disclosing, Reporting, and Blocking," May 2021, [https://info.thorn.org/hubfs/Research/Responding%20to%20Online%20Threats\\_2021-Full-Report.pdf?utm\\_campaign=H2D%20report&utm\\_source=website](https://info.thorn.org/hubfs/Research/Responding%20to%20Online%20Threats_2021-Full-Report.pdf?utm_campaign=H2D%20report&utm_source=website).
- 11 Kathryn C. Montgomery and Jeff Chester, "Data Protection for Youth in the Digital Age: Developing a Rights-Based Global Framework," *European Data Protection Law Review* 1, no. 4 (2015): 277–91. <https://doi.org/10.21552/edpl/2015/4/6>.
- 12 CIGI, "The International Grand Committee on Disinformation and 'Fake News,'" <https://www.cigionline.org/igc>; Shannon Bond, "Civil Rights Groups Say If Facebook Won't Act On Election Misinformation, They Will," NPR, 25 Sept. 2020, <https://www.npr.org/2020/09/25/916782712/civil-rights-groups-say-if-facebook-wont-act-on-election-misinformation-they-wil>; Shirin Ghaffary, "Civil Rights Leaders are Still Fed Up with Facebook Over Hate Speech," *Vox*, 7 July 2020, <https://www.vox.com/recode/2020/7/7/21316681/facebook-mark-zuckerberg-civil-rights-hate-speech-stop-hate-for-profit>; Richard Nieva, "YouTube Criticized for Misinformation Over Videos Falsely Claiming Trump Victory," *CNET*, 5 Nov. 2020, <https://www.cnet.com/news/youtube-criticized-for-misinformation-over-videos-falsely-claiming-trump-victory/>.
- 13 Federal Trade Commission, "Lina M. Khan Sworn in as Chair of the FTC," 15 June 2021, <https://www.ftc.gov/news-events/news/press-releases/2021/06/lina-m-khan-sworn-chair-ftc>; "House Lawmakers Release Anti-Monopoly Agenda for 'A Stronger Online Economy: Opportunity, Innovation, Choice,'" 11 June 2021, <https://cicilline.house.gov/press-release/house-lawmakers-release-anti-monopoly-agenda-stronger-online-economy-opportunity>.
- 14 "Senators Markey and Cassidy Propose Bipartisan Bill to Update Children's Online Privacy Rules," 24 June 2021, <https://www.markey.senate.gov/news/press-releases/senators-markey-and-cassidy-propose-bipartisan-bill-to-update-childrens-online-privacy-rules>; "Rep. Castor Reintroduces Landmark Kids PRIVCY Act to Strengthen COPPA, Keep Children Safe Online," 29 July, 2021, <https://castor.house.gov/news/documentsingle.aspx?DocumentID=403677>; "Blackburn & Blumenthal Introduce Comprehensive Kids' Online Safety Legislation," 16 Feb. 2022, <https://www.blackburn.senate.gov/2022/2/blackburn-blumenthal-introduce-comprehensive-kids-online-safety-legislation>.

- 15 “AB-2273 The California Age-Appropriate Design Code Act. (2021-2022),” [https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill\\_id=202120220AB2273](https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202120220AB2273).
- 16 United Kingdom, Information Commissioner’s Office, “Introduction to the Age Appropriate Design Code,” <https://ico.org.uk/for-organisations/guide-to-data-protection/ico-codes-of-practice/age-appropriate-design-code/>; 5Rights Foundation, “The Digital Services Act Must Deliver for Children,” <https://5rightsfoundation.com/in-action/the-digital-services-act-must-deliver-for-children.html>.
- 17 We spearheaded a national campaign that led to the 1998 passage of the federal Children’s Online Privacy Protection Act (COPPA). See Montgomery, *Generation Digital*, pp. 67-106. We have also conducted research on other arenas that are part of the broad, data-driven digital tech sector, principally our 2016 report on health wearables, and several papers on how the same digital techniques and strategies that brands use to influence consumers have been adopted and adapted by political operatives and are now central to contemporary campaigns for both candidates and issues. See, for example, Jeff Chester, Kathryn C. Montgomery, Katharina Kopp, “‘Big Food’ and ‘Big Data’ Online Platforms Fueling Youth Obesity Crisis as Coronavirus Pandemic Rages,” 12 May 2021, <https://www.democraticmedia.org/article/big-food-and-big-data-online-platforms-fueling-youth-obesity-crisis-coronavirus-pandemic-1>; Kathryn Montgomery, Jeff Chester, and Katharina Kopp, “Health Wearable Devices Pose New Consumer and Privacy Risks,” Center for Digital Democracy, 29 Aug. 2017, <https://www.democraticmedia.org/blog/health-wearable-devices-pose-new-consumer-and-privacy-risks>.
- 18 The research for this report has included a variety of sources: interviews, analyses of industry reports and documents, and examination of tech company policies and government initiatives.
- 19 Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media* (New Haven, CT: Yale University Press, 2018), p. 124.
- 20 Gillespie, *Custodians of the Internet*, pp. 30-31.
- 21 Gillespie, *Custodians of the Internet*, p.136.
- 22 The people who do this work often have to endure terrible working conditions, suffering frequently from PTSD and other serious afflictions, as scholar Sarah Roberts has documented in her book, *Behind the Screen: Content Moderation in the Shadows of Social Media* (New Haven, CT: Yale University Press, 2019).
- 23 Paul M. Barrett, “Who Moderates the Social Media Giants? A Call to End Outsourcing,” Stern Center for Business and Human Rights, June 2020, [https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/5ed9854bf618c710cb55be98/1591313740497/NYU+Content+Moderation+Report\\_June+8+2020.pdf](https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/5ed9854bf618c710cb55be98/1591313740497/NYU+Content+Moderation+Report_June+8+2020.pdf).
- 24 Barrett, “Who Moderates the Social Media Giants?”
- 25 Gillespie, *Custodians of the Internet*, p. 78
- 26 Canadian Centre for Child Protection, “Reviewing Child Sexual Abuse Material Reporting Functions on Popular Platforms,” <https://protectchildren.ca/en/resources-research/csam-reporting-platforms/>.
- 27 Catherine Shu, “Facebook Says it Removed 8.7M Child Exploitation Posts with New Machine Learning Tech,” TechCrunch, 24 Oct. 2018, <https://techcrunch.com/2018/10/24/facebook-says-it-removed-8-7m-child-exploitation-posts-with-new-machine-learning-tech/>; “18 U.S. Code § 2258A - Reporting Requirements of Providers,” Cornell Law School Legal Information Institute, <https://www.law.cornell.edu/us-code/text/18/2258A>.
- 28 Microsoft, “Help Stop the Spread of Child Exploitation,” <https://www.microsoft.com/en-us/photodna>.
- 29 Michael Salter and Elly Hanson, “I Need You To Understand How Pervasive This Issue Is: User Efforts to Regulate Child Sexual Offending on Social Media,” in Michael Salter and Elly Hanson, eds, *The Emerald International Handbook of Technology Facilitated Violence and Abuse*, June 2021, p. 732, <https://www.emerald.com/insight/publication/doi/10.1108/9781839828485>.
- 30 Salter and Hanson, “I Need You To Understand How Pervasive This Issue Is,” p. 732. Thorn also has a software system, called Safer, designed for smaller to medium size companies. Thorn, “The Road to Safer: Equipping Industry to End CSAM,” 21 July 2020, <https://www.thorn.org/blog/announcing-safer-built-by-thorn-eliminate-csam/>.
- 31 World Federation of Advertisers, “GARM Aggregated Measurement Report – November 2021,” <https://wfanet.org/knowledge/item/2021/11/30/GARM-Aggregated-Measurement-Report-November-2021-3>; TikTok, “Community Guidelines Enforcement Report: April 1, 2021 – June 30, 2021,” 13 Oct. 2021 <https://www.tiktok.com/transparency/en-us/community-guidelines-enforcement-2021-2/>; Guy Rosen, “Community Standards Enforcement Report, Second Quarter 2021,” Meta, 18 Aug. 2021, <https://about.fb.com/news/2021/08/community-standards-enforcement-report-q2-2021/>.
- 32 World Federation of Advertisers, “GARM Aggregated Measurement Report – November 2021.”
- 33 See, for example, Google, “About Integral Ad Science’s Media Quality Verification,” <https://support.google.com/displayvideo/answer/3297897?hl=en>; Meta, “How do Facebook’s Third-party Viewability Verification Integrations Work?” <https://www.facebook.com/business/help/313903878988514>; Integral Ad Science, “TikTok Brand Safety Solution: Make Every Impression Count,” [https://go.integralads.com/rs/469-VBI-606/images/IAS\\_TikTok\\_Brand\\_Safety\\_Solution\\_One\\_Sheet.pdf](https://go.integralads.com/rs/469-VBI-606/images/IAS_TikTok_Brand_Safety_Solution_One_Sheet.pdf).

- 34 Mark Bridge, "Child Abuse on YouTube," *The Times*, 18 Nov. 2017, <https://www.thetimes.co.uk/article/child-abuse-on-youtube-q3x9zfkch>; Alexi Mostrous, "Big Brands Fund Terror Through Online Adverts," *The Times*, 9 Feb. 2017, <https://www.thetimes.co.uk/article/big-brands-fund-terror-knnxfgb98>; "UK Government Pulls Ads from YouTube in Extremist Content Row," *New Scientist*, 17 Mar. 2017, <https://www.bloomberg.com/news/articles/2017-03-17/u-k-pulls-ads-from-youtube-citing-failures-in-policing-content?sref=QDmhoVl8>. Marketers were able to leverage the resulting controversy generated by the press, which they fueled by relatively short-lived and well-publicized boycotts, to force Google and Facebook to adopt a number of changes designed to give advertisers greater influence and oversight. Meta, "Introducing a Brand Safety Certification for Facebook Marketing Partners," *Meta for Business*, 27 May 2020, <https://www.facebook.com/business/marketing-partners/partner-news/introducing-a-brand-safety-certification-for-facebook-marketing-partners>; Meta, "Accountability and New Choices for Marketers," *Meta for Business*, 10 Feb. 2017, <https://www.facebook.com/business/news/accountability-and-new-choices-for-marketers>; George P. Slefo, "YouTube to Offer Third-Party Brand Safety Tools Following Revolt by Marketers," *Ad Age*, 3 Apr. 2017, <https://adage.com/article/digital/youtube-offer-party-brand-safety-tools/308525>; Allison Schiff, "YouTube Expands Its Measurement Program with 5 New Partners, Bumps Heads with OpenSlate," *AdExchanger*, 20 Apr. 2020, <https://www.adexchanger.com/analytics/youtube-expands-its-measurement-program-with-5-new-partners-bumps-heads-with-existing-partner-openslate/>.
- 35 IAB, "Transparency Center Evolves to Further Improve Supply Chain Safety," 2 Dec. 2021 <https://iabtechlab.com/blog/transparency-center-evolves-to-further-improve-supply-chain-safety/>.
- 36 World Federation of Advertisers, "Marketing Leaders Take Action on Harmful Online Content," 22 Jan. 2020, <https://wfanet.org/knowledge/item/2020/01/22/Marketing-leaders-take-action-on-harmful-online-content>.
- 37 Global Alliance for Responsible Media, "GARM: Brand Safety Floor + Suitability Framework," <https://wfanet.org/library/download/urn:uuid:7d484745-41cd-4cce-a1b9-a1b4e30928ea/garm+brand+safety+floor+suitability+framework+23+sept.pdf>.
- 38 The other categories defined by GARM involve "Hate speech and acts of aggression," defined as "behavior or content that incites hatred, promotes violence, vilifies, or dehumanizes groups or individuals based on race, ethnicity, gender, sexual orientation, gender identity, age, ability, nationality, religion, caste, victims and survivors of violent acts and their kin, immigration status, or serious disease suffers"; "Arms and Ammunition; Death, Injury or Military Conflict; Online Piracy; Obscenity and Profanity, including language, gestures, and explicitly gory, graphic or repulsive content intended to shock and disgust; Illegal Drugs/Tobacco/ecigarettes/Vaping/Alcohol (where it is illegal to do so); Spam or Harmful Content (identified as "malware and phishing); Terrorism; and Debated Sensitive Social Issues." Global Alliance for Responsible Media, "GARM: Brand Safety Floor + Suitability Framework."
- 39 Global Alliance for Responsible Media, "GARM: Brand Safety Floor + Suitability Framework." This more flexible schema is open to further alterations. The Interactive Advertising Bureau (IAB), which is the key online ad technology and marketing trade group, has set up a "brand suitability" testing environment, which is available through an industry open-source-like system, so that platforms, advertisers and others can assess for themselves how well GARM's framework operates with "actual content." "IABTechLab/Brand-Suitability-Test-Benchmarks," <https://github.com/IABTechLab/Brand-Suitability-Test-Benchmarks>. (In addition to exploring brand suitability, the IAB's Tech Lab has working groups on several key issues, including one on "taxonomy and mapping," to help develop efficient ways to classify content for marketing purposes).
- 40 DoubleVerify, "DoubleVerify Launches Quality Coverage Across Facebook Stories, Offering Greater Transparency and Control within the Facebook Platform," 19 Dec. 2018, <https://www.doubleverify.com/newsroom/doubleverify-launches-quality-coverage-across-facebook-stories-offering-greater-transparency-and-control-within-the-facebook-platform/>; Integral Ad Science, "Google Selects IAS as Brand Safety and Viewability Preferred Measurement Partner," 12 July 2018, <https://integralads.com/news/google-selects-ias-brand-safety-viewability-preferred-measurement-partner/>.
- 41 IAS has "deep integrations with all the major advertising and technology platforms," including Amazon, Facebook, Google/YouTube, Instagram, Snap, Twitter and the Trade Desk. DV has similar partnerships, including Google/YouTube, Facebook, Pinterest, SNAP, Twitter, Amazon Advertising and the Trade Desk. Tony Rifilato, "IAS Goes Public In Bid To Retain Competitive Edge," *AdExchanger*, 30 June 2021, <https://www.adexchanger.com/platforms/ias-goes-public-in-bid-to-retain-competitive-edge/>; "DoubleVerify Announces Integration with YouTube," 12 Jan. 2017, <https://doubleverify.com/newsroom/doubleverify-announces-integration-with-youtube/>; "DoubleVerify Launches Quality Coverage Across Facebook Stories, Offering Greater Transparency and Control Within the Facebook Platform," 19 Dec. 2018, <https://doubleverify.com/newsroom/doubleverify-launches-quality-coverage-across-facebook-stories-offering-greater-transparency-and-control-within-the-facebook-platform/>.
- 42 Integral Ad Science, "Unlock the Power of Context to Protect and Grow Your Brand," <https://integralads.com/solutions/brand-safety-suitability/>; "DoubleVerify Launches Semantic Science Division to Bolster Brand Safety & Performance Solutions," 18 Sept. 2019, <https://doubleverify.com/doubleverify-launches-semantic-science-division-to-bolster-brand-safety-performance-solutions/>.
- 43 Integral Ad Science, "Form 10-K filing, United States Securities and Exchange Commission," 2021, <https://investors.integralads.com/node/7376/html>; Integral Ad Science, "A Global Leader in Digital Media Quality," <https://integralads.com/about-ias/>.
- 44 DoubleVerify, "Investor Presentation," July 2021, [https://d1io3yog0oux5.cloudfront.net/\\_3d94986c09acc81382edadc827c84397/doubleverify/db/1086/9754/pdf/Investor+Presentation+Interim+-+July+2021+%281%29.pdf](https://d1io3yog0oux5.cloudfront.net/_3d94986c09acc81382edadc827c84397/doubleverify/db/1086/9754/pdf/Investor+Presentation+Interim+-+July+2021+%281%29.pdf).

- 45 Tim Mahlman, "Despite Challenges, Programmatic is Evolving to Deliver on Brand Safety," AdExchanger, 14 Aug. 2017, <https://adexchanger.com/data-driven-thinking/despite-challenges-programmatic-evolving-deliver-brand-safety/>. For example, IAS has established eight standard content categories that have been scored for brand-safety purposes, involving adult content, alcohol, gambling, hate speech, illegal downloads, illegal drugs, offensive language, and violence. Several hundred other content categories are also available that can either be avoided or targeted for ads, based on the "advertiser's specific risk tolerance." IAS and DV, as well as others, provide what is known as "pre-bid" services that enable ads to be targeted in environments deemed to be safe and suitable. See, for example, DoubleVerify, "Bid Intelligently with DV's Privacy-Friendly Programmatic Solutions," [https://doubleverify.com/wp-content/uploads/2021/10/DV\\_OneSheet\\_ProgrammaticSolutions\\_Standard.pdf](https://doubleverify.com/wp-content/uploads/2021/10/DV_OneSheet_ProgrammaticSolutions_Standard.pdf).
- 46 DoubleVerify, "Investor Presentation."
- 47 OpenSlate, "Kids Content & YouTube Brand Safety," 3 Apr. 2019, <https://www.openslate.com/blog-posts/content-matters-kids-content-youtube-brand-safety/>; OpenSlate, "Content Matters: OpenSlate Launches Facebook Brand Safety Solution," 24 Jan. 2019, <https://www.openslate.com/solutions/>; "OpenSlate and GroupM Partner to Expand Brand Suitability on YouTube," GlobeNewsWire, 26 June 2018, <https://www.openslate.com/blogs/openslate-launches-facebook-brand-safety-solution/>; OpenSlate, "Content Matters: Kids Content & YouTube Brand Safety," 3 Apr. 2019, <https://www.openslate.com/blog-posts/content-matters-kids-content-youtube-brand-safety/>. DoubleVerify announced plans to acquire OpenSlate late last year. OpenSlate, "OpenSlate to Be Acquired by DoubleVerify," 9 Nov. 2021, <https://www.openslate.com/blog-posts/openslate-to-be-acquired-by-doubleverify/>.
- 48 Picalate, "Safe & Trustworthy Ad Supply Chain," <https://www.picalate.com/use-cases/>; Picalate, "Publisher Trust Index: Methodology," <https://www.picalate.com/publisher-trust-index-methodology>.
- 49 IPG Mediabrands, "Mediabrand Study Shows Advertisers Risk Getting Ensnared in Accelerating Misinformation Across Media & Social Media Platforms," 27 Sept. 2021, <https://www.ipgmediabrands.com/mediabrands-study-shows-advertisers-risk-getting-ensnared-in-accelerating-misinformation-across-media-social-media-platforms/>.
- 50 Mozilla, "YouTube Regrets: A Crowdsourced Investigation into YouTube's Recommendation Algorithm" July 2021, [https://assets.mofoprod.net/network/documents/Mozilla\\_YouTube\\_Regrets\\_Report.pdf](https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf).
- 51 Chris Sutcliffe, "Big Brands Pull Ads as YouTube Battles Conspiracy and Misinformation Videos," The Drum, 14 Feb. 2022, <https://www.thedrum.com/news/2022/02/14/big-brands-pull-ads-youtube-battles-conspiracy-and-misinformation-videos>.
- 52 ID Comms, "ID Comms Global Transparency Report 2021," <https://content.idcomms.com/download-the-id-comms-2021-global-media-transparency-report> (registration required).
- 53 Global Alliance for Responsible Media, "GARM Aggregated Measurement Report," vol. 1, Apr. 2021, <https://www.isanet.org.in/crm/articlepdf/GARM%20Aggregated%20Measurement%20Report%20Vol%201.pdf>.
- 54 Global Alliance for Responsible Media, "GARM Aggregated Measurement Report," vol. 1.
- 55 Garrett Sloane and Mike Juang, "Brand Safety Report Reveals the Dangers for Advertisers on Platforms Like Facebook and YouTube," *Ad Age*, 20 Apr. 2021, <https://adage.com/article/digital/brand-safety-report-reveals-dangers-advertisers-platforms-facebook-and-youtube/2329431>.
- 56 Digital Trust & Safety Partnership, "The Safe Framework: Tailoring a Proportionate Approach to Assessing Digital Trust & Safety," Dec. 2021, <https://dtspartnership.org/>.
- 57 Facebook, "What is the Facebook Safety Advisory Board and What Does this Board Do?" <https://www.facebook.com/help/222332597793306>; Twitch, "Introducing the Twitch Safety Advisory Council," *Twitch Blog*, 14 May 2020, <https://blog.twitch.tv/en/2020/05/14/introducing-the-twitch-safety-advisory-council/>; TikTok, "Safety Partners," <https://www.tiktok.com/safety/en-us/safety-partners/>.
- 58 Associated Press, "How Sweeping E.U. Rules Would Curb Tech Companies," NBC News, 20 Jan. 2022, <https://www.nbcnews.com/tech/tech-news/sweeping-eu-rules-curb-tech-companies-rcna12868>; Natasha Lomas "UK Now Expects Compliance with Children's Privacy Design Code," TechCrunch, 1 Sept. 2021, <https://techcrunch.com/2021/09/01/uk-now-expects-compliance-with-its-child-privacy-design-code/>.
- 59 Sara Lebow, "More US Children Consume YouTube Videos Than Any Other Type of Media," eMarketer Insider Intelligence, 15 Oct. 2021, personal copy.
- 60 Sheera Frenkel and Cecilia Kang, *An Ugly Truth: Inside Facebook's Battle for Domination* (New York: Harper, 2021), pp. 227-231.
- 61 Zak Doffman, "Instagram The Worst As Social Media Slammed As 'A Gateway For Child Abuse,'" *Forbes*, 1 Mar 2029, <https://www.forbes.com/sites/zakdoffman/2019/03/01/surge-of-200-in-use-of-instagram-to-target-and-abuse-children/?sh=8b97c7159c4e>.
- 62 Andrew Hutchinson, "New Report Finds Snapchat, TikTok Outpace Instagram in Teen Popularity," Social Media Today, 7 Oct. 2020, <https://www.socialmediatoday.com/news/new-report-finds-snapchat-tiktok-outpace-instagram-in-teen-popularity/586614/>.
- 63 Omnicore, "Instagram by the Numbers: Stats, Demographics & Fun Facts," 27 Feb. 2022, <https://www.omnicoreagency.com/instagram-statistics/>; Salvador Rodriguez, "TikTok Usage Surpassed Instagram This Year Among Kids Aged 12 to 17, Forrester Survey Says," CNBC, 18 Nov. 2021, <https://www.cnbc.com/2021/11/18/tiktok-usage-topped-instagram-in-2021-among-kids-12-to-17-forrester.html>; Piper Sandler, "Taking Stock with Teens," Spring 2022, <https://www.pipersandler.com/1col.aspx?id=6216>; Katherine Schaeffer, "7 Facts about Americans and Instagram," Pew Research Center, 7 Oct. 2021, <https://www.pewresearch.org/fact-tank/2021/10/07/7-facts-about-americans-and-instagram/>.
- 64 Instagram, "When Does My Instagram Story Disappear?" <https://help.instagram.com/1729008150678239>.



- 65 Meta, "Facebook Community Standards," Transparency Center, <https://transparency.fb.com/policies/community-standards/>.
- 66 Doffman, "Instagram The Worst As Social Media Slammed As 'A Gateway For Child Abuse.'"
- 67 Ryan Mac and Craig Silverman, "Facebook Is Building An Instagram For Kids Under The Age Of 13," Buzz Feed News, 18 Mar. 2021, <https://www.buzzfeednews.com/article/ryanmac/facebook-instagram-for-children-under-13>.
- 68 Alex Hern, "Social Media Giants Increase Global Child Safety After UK Regulations Introduced," *The Guardian*, 5 Sept. 2021, <https://www.theguardian.com/media/2021/sep/05/social-media-giants-increase-global-child-safety-after-uk-regulations-introduced>.
- 69 Thorn, "Responding to Online Threats: Minors' Perspectives on Disclosing, Reporting, and Blocking."
- 70 Color of Change, "Stop Hate for Profit: Calling on Facebook Corporate Advertisers to Pause Ads for July 2020," <https://colorofchange.org/stop-hate-for-profit/>.
- 71 Meta, "Content Inventory Table: Facebook Placements," Meta Business Help Center, <https://www.facebook.com/business/help/468154293719742?id=1769156093197771>; Meta, "Brand Safety Controls," Meta Business Help Center, [https://www.facebook.com/business/help/1926878614264962?id=1769156093197771&locale=en\\_US](https://www.facebook.com/business/help/1926878614264962?id=1769156093197771&locale=en_US); Meta, "About Inventory Filter," Meta Business Help Center, <https://www.facebook.com/business/help/3001448133206080?id=1769156093197771>.
- 72 "Meta said the Facebook News Feed controls were effective 99% of the time, keeping brands away from posts related to 'tragedy and conflict.' It was 94% effective at keeping ads away from news and political adjacency...." Garrett Sloane, "Facebook Expands Brand Safety Controls in News Feed and Instagram," *Ad Age*, 18 Nov. 2021, <https://adage.com/article/digital-marketing-ad-tech-news/facebook-expands-brand-safety-controls-news-feed-and-instagram/2380971>.
- 73 Allison Schiff, "OpenSlate Overhauls Its Content Rating Tool And Expands From YouTube To Facebook," *AdExchanger*, 6 Feb. 2020, <https://www.adexchanger.com/digital-tv/openslate-overhauls-its-content-rating-tool-and-expands-from-youtube-to-facebook/>; "OpenSlate Launches Universal SlateScore Metric for Digital Video," *GlobeNewswire*, 6 Feb. 2020, <https://www.globenewswire.com/en/news-release/2020/02/06/1981343/0/en/OpenSlate-Launches-Universal-SlateScore-Metric-for-Digital-Video.html>.
- 74 As the company explained, it will focus on "accelerating our integrity and privacy work to ensure the safest possible experience for teens, and building a version of Instagram that allows people under the age of 13 to safely use Instagram for the first time." Mac and Silverman, "Facebook Is Building An Instagram For Kids Under The Age Of 13."
- 75 Terry Collins, "Instagram's New Safety, Privacy Tools for Teens Includes Adult-minor Restrictions in DMs," *USA Today*, 16 Mar. 2021, [https://www.usatoday.com/story/tech/2021/03/16/instagram-privacy-safety-updates-facebook/4705431001/?mkt\\_tok=MTM4LUVaTS0wN-DIAAAF74CwB5Z1ns\\_c3AUAAPTTSWjmd8Pbf0AWRcXEpsMOXe6R-25M66lleu-rdqN5EubcqBtTI0MGMLScpCg19onJYj2Jh8Nw5XtfYV-jfJVCexxNe](https://www.usatoday.com/story/tech/2021/03/16/instagram-privacy-safety-updates-facebook/4705431001/?mkt_tok=MTM4LUVaTS0wN-DIAAAF74CwB5Z1ns_c3AUAAPTTSWjmd8Pbf0AWRcXEpsMOXe6R-25M66lleu-rdqN5EubcqBtTI0MGMLScpCg19onJYj2Jh8Nw5XtfYV-jfJVCexxNe).
- 76 Geoff Weiss, "Instagram Announces New Teen Safety Features In DMs Amid Gradual Move To Encryption," *Tubefilter*, 16 Mar. 2021, <https://www.tubefilter.com/2021/03/16/instagram-announces-new-safety-features-for-teens/>; Child Mind Institute, "A Beacon for Children's Mental Health: Guiding Families in a Time of Crisis," 2020 annual report, <https://childmind.org/wp-content/uploads/ChildMindInstitute2020AnnualReport.pdf>.
- 77 Sarah Perez, "Consumer Groups and Child Development Experts Petition Facebook to Drop 'Instagram for Kids' Plan," *TechCrunch*, 15 Apr. 2021, <https://techcrunch.com/2021/04/15/consumer-groups-and-child-development-experts-petition-facebook-to-drop-instagram-for-kids-plan/>.
- 78 Adam Satariano and Ryan Mac, "Facebook is Delaying an Instagram App for Users Under the Age of 13," *Seattle Times*, 27 Sept. 2021, <https://www.seattletimes.com/business/facebook-is-delaying-an-instagram-app-for-users-under-the-age-of-13/>.
- 79 Sarah Perez, "Instagram Announces Plans for Parental Controls and Other Safety Features Ahead of Congressional Hearing," *TechCrunch*, 7 Dec. 2021, <https://techcrunch.com/2021/12/07/instagram-announces-plans-for-parental-controls-and-other-safety-features-ahead-of-congressional-hearing/>.
- 80 Perez, "Instagram Announces Plans for Parental Controls and Other Safety Features Ahead of Congressional Hearing."
- 81 Perez, "Instagram Announces Plans for Parental Controls and Other Safety Features Ahead of Congressional Hearing."
- 82 Instagram, "New Tools and Resources for Parents and Teens in VR and on Instagram," 14 June 2022, <https://about.instagram.com/blog/announcements/tools-and-resources-for-parents-and-teens-in-vr-and-on-instagram>.
- 83 Kate Kaye, "Why Facebook's Limits on Teen Targeting Are All Part of its Algorithmic Ad Playbook," *Digiday*, 4 Aug. 2021, <https://digiday.com/marketing/why-facebooks-limits-on-teen-targeting-are-all-part-of-its-algorithmic-ad-playbook/>.
- 84 Kaye, "Why Facebook's Limits on Teen Targeting Are All Part of its Algorithmic Ad Playbook."
- 85 Andrei Ene, "Facebook Messenger and Instagram Will Not Support End-to-End Encryption Any Time Soon," *TechTheLead*, 3 May 2021, <https://techthelead.com/facebook-messenger-and-instagram-will-not-support-end-to-end-encryption-any-time-soon/>.
- 86 Gail Kent, "Messenger Policy Workshop: Future of Private Messaging," Meta, 30 Apr. 2021, <https://about.fb.com/news/2021/04/messenger-policy-workshop-future-of-private-messaging/>.
- 87 Andrew Hutchinson, "Meta Moves to the Next Stage of its Full Messaging Integration Plan," *Social Media Today*, 13 Sept. 2022, <https://www.socialmediatoday.com/news/meta-moves-to-the-next-stage-of-its-full-messaging-integration-plan/631806/>.

- 88 Emily McCormick, "Snapchat Widens Lead Over TikTok, Instagram as Teens' Favorite Social Media App: Piper Sandler," Yahoo! Finance, 5 Oct. 2021, <https://finance.yahoo.com/news/snapchat-widens-lead-over-tik-tok-instagram-as-teens-favorite-social-media-app-piper-sandler-145605293.html>; "Piper Sandler Completes 43rd Semi-Annual Generation Z Survey of 7,100 U.S. Teens," Business Wire, 6 Apr. 2022, <https://www.yahoo.com/now/piper-sandler-completes-43rd-semi-124900072.html>.
- 89 Christine Lagorio-Chafkin, "Don't Credit Sexting: How Snapchat Actually Took Off," Inc., 5 Jan. 2021, <https://www.inc.com/christine-lagorio/real-origins-of-snapchat-growth.html>.
- 90 Snapchat, "Snapchat Ads Targeting," <https://forbusiness.snapchat.com/advertising/targeting>.
- 91 WARC, "What We Know About Marketing on Snapchat," WARC Best Practice, Jan. 2022, <https://www.warc.com/content/article/bestprac/what-we-know-about-marketing-on-snapchat/109912>; Snap, Inc., "Announcing New Policies for Snap's Developer Platform," 17 Mar. 2022, <https://snap.com/en-US/safety-and-impact>.
- 92 Snap, "Report Abuse on Snapchat," Snapchat Support, <https://support.snapchat.com/en-US/a/report-abuse-in-app>; Snap, "Safety Center," <https://snap.com/en-US/safety/safety-center>.
- 93 Snap, Inc., "Meet Our Head of Global Platform Safety," 25 Jan. 2022, <https://snap.com/en-US/safety-and-impact/post/meet-our-head-of-global-platform-safety>.
- 94 Snap, Inc., "Meet Our Head of Global Platform Safety."
- 95 Canadian Centre for Child Protection, "Reviewing Child Sexual Abuse Material Reporting Functions on Popular Platforms: Executive Summary," [https://www.protectchildren.ca/pdfs/C3P\\_ReviewingCSAMMaterialReporting\\_Summary\\_en.pdf](https://www.protectchildren.ca/pdfs/C3P_ReviewingCSAMMaterialReporting_Summary_en.pdf).
- 96 Zak Doffman, "Snapchat Has Become A 'Haven For Child Abuse' With Its 'Self-Destructing Messages,'" *Forbes*, 26 May 2019, <https://www.forbes.com/sites/zakdoffman/2019/05/26/snapchats-self-destructing-messages-have-created-a-haven-for-child-abuse/?sh=49f3087d399a>; Rosamund Urwin and Tom Calver, "Predators Flock to Teen App Snapchat, a Safe Space for Child Abuse," *The Sunday Times*, 26 May 2019, <https://www.thetimes.co.uk/article/predators-flock-to-snapchat-a-safe-space-for-child-abuse-9ztwf0tf9>.
- 97 Doffman, "Snapchat Has Become A 'Haven For Child Abuse' With Its 'Self-Destructing Messages.'"
- 98 Snap, Inc., "Investing in and Expanding our Law Enforcement Operations," 2 Dec. 2021, <https://snap.com/en-US/safety-and-impact/post/investing-in-and-expanding-our-law-enforcement-operations>.
- 99 Snap, Inc., "Transparency Report: January 1, 2021 – June 30, 2021," <https://snap.com/en-US/privacy/transparency>.
- 100 "Testimony of Jennifer Stout, Vice President of Global Public Policy, Snap Inc., Hearing before the United States Senate Committee on Science, Commerce, and Transportation, Subcommittee on Consumer Protection, Product Safety, and Data Security," 26 Oct. 2021, [https://assets.ctfassets.net/gqgsr8avay9x/1cR28yqyK12gRyp4D-vAdFK/05056b6b21d5ed8be4d71e19299dec99/Testimony\\_of\\_Jennifer\\_Stout\\_VP\\_of\\_Global\\_Public\\_Policy\\_\\_Snap\\_Inc.pdf](https://assets.ctfassets.net/gqgsr8avay9x/1cR28yqyK12gRyp4D-vAdFK/05056b6b21d5ed8be4d71e19299dec99/Testimony_of_Jennifer_Stout_VP_of_Global_Public_Policy__Snap_Inc.pdf).
- 101 Snapchat, "Introducing Here for You," 11 Feb. 2020, <https://newsroom.snap.com/en-GB/here-for-you/>; Andrew Hutchinson, "Snapchat Launches New 'Here For You' Mental Health Resource Ahead of Time in Response to COVID-19," *Social Media Today*, 20 Mar. 2020, <https://www.socialmediatoday.com/news/snapchat-launches-new-here-for-you-mental-health-resource-ahead-of-time-i/574507/>.
- 102 "Boys Aggressively Targeted on Instagram and Snapchat, Analysis of Cybertip.ca Data Shows," *Cybertip.ca*, 4 Aug. 2022, <https://www.cybertip.ca/en/campaigns-and-media/news-releases/2022/sexortion-data-analysis/>.
- 103 Snap Inc., "Introducing Family Center on Snapchat," 9 Aug. 2022, <https://snap.com/en-US/safety-and-impact/post/family-center>.
- 104 Faez Samadi, "Snap Inc Partners with Integral Ad Science Over Brand Safety," *Campaign*, 30 May 2017, <https://www.campaignlive.com/article/snap-inc-partners-integral-ad-science-brand-safety/1435022>; Tobi Elkin, "DoubleVerify, IAS, And Moat Join Snapchat Brand Safety Coalition," *Media Post*, 30 May 2017, <https://www.mediapost.com/publications/article/302020/doubleverify-ias-and-moat-join-snapchat-brand-sa.html>; Snap, "Brand Safety," Snapchat Support, <https://support.snapchat.com/en-US/a/brand-safety>; Snap, Inc., "Snapchat Safety Center," <https://www.snap.com/en-US/safety/safety-center>.
- 105 Snap, Inc., "Safety Resources and Support," <https://snap.com/en-US/safety/safety-resources>.
- 106 Rights Foundation, "Pathways: How Digital Design Puts Children at Risk," July 2021, <https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf>.
- 107 "What We Know about Marketing on TikTok," WARC Best Practice, Mar. 2020, <https://www.warc.com/content/paywall/article/bestprac/what-we-know-about-marketing-on-tiktok/131979>, <https://www.warc.com/content/article/bestprac/what-we-know-about-marketing-on-tiktok/131979>.
- 108 eMarketer, "TikTok Users, By Age: US2022," <https://forecasts-na1.emarketer.com/5e4c6b5ec56a401090de843e/60883a7715ad-620b84502a14> (subscription required); Debra Aho Williamson, "US Social Video Usage 2022," eMarketer, 7 Mar. 2022, <https://content-na1.emarketer.com/us-social-video-usage-2022>; Mark Sweney, "TikTok Reports \$1bn Turnover Across International Markets," *The Guardian*, 4 Oct. 2022, <https://www.theguardian.com/technology/2022/oct/04/tiktok-reports-1bn-turnover-across-international-markets>.
- 109 Ryan Holmes, "Is TikTok a Time Bomb?" *Fast Company*, 28 Aug. 2019, <https://www.fastcompany.com/90395898/is-tiktok-a-time-bomb>.

- 110 Sarah Perez, "Kids Now Spend Nearly as Much Time Watching TikTok as YouTube in US, UK and Spain," TechCrunch, 4 June 2020, <https://techcrunch.com/2020/06/04/kids-now-spend-nearly-as-much-time-watching-tiktok-as-youtube-in-u-s-u-k-and-spain/>; "Content Discovery and Creation Powered by AI and ML – Building the Future," Asia Tech Daily, 7 Dec. 2018, <https://www.asiatechdaily.com/content-discovery-and-creation-powered-by-ai-and-ml-building-the-future/>; TikTok, "Welcome to TikTok Ads Manager!" <https://ads.tiktok.com/help/article?aid=6667447877242978309>.
- 111 TikTok, "Smart Video," <https://ads.tiktok.com/help/article?aid=6760164394144890886>.
- 112 "With 'Automated Creative Optimization,' you can upload images or videos, write some ad text, and select your call-to-action (CTA) buttons. Our system will then automatically combine your creative assets into multiple ads for your campaign. These ads get continuously explored, evaluated and optimized to find the optimal combination of variables. The system will then present the best creative to your target audience based on the tested combinations." TikTok, "Automated Creative Optimization," <https://ads.tiktok.com/help/article?aid=6670043695674294277>.
- 113 TikTok, "Third Party Tracking," <https://ads.tiktok.com/help/article?aid=6669666107177041925>; TikTok "TikTok Ads Attribution," <https://ads.tiktok.com/help/article?aid=6704457722424721413>; Conviva, "TikTok Analytics," <https://www.prnewswire.com/news-releases/conviva-launches-tiktok-analytics-becoming-the-first-to-enable-brands-to-measure-collective-social-media-success-301062992.html>.
- 114 For examples of TikTok marketing, see TikTok, "Sour Patch Kids: Driving Strong Consideration and Efficient CTR with Display Cards," <https://www.tiktok.com/business/en-US/inspiration/sour-patch-kids-568>; TikTok, "Mars: Increasing Awareness and Driving App Installs of Mars Wrigley's Virtual Trick-or-Treating App," <https://www.tiktok.com/business/en-US/inspiration/mars-331>; TikTok, "McDonald's: Enlisting a Star-studded Cast of Creators to Build Brand Love Around a New Beat," <https://www.tiktok.com/business/en-US/inspiration/mc-donald's-401>.
- 115 Elizabeth Culliford, "TikTok Reveals First Members of New U.S. Content-moderation Committee," Reuters, 18 Mar. 2020, <https://www.reuters.com/article/us-tiktok-content/tiktok-reveals-first-members-of-new-u-s-content-moderation-committee-idUSKBN2152BJ>; Greg Roumeliotis, Yingzhi Yang, Echo Wang, and Alexandra Alper, "Exclusive: U.S. Opens National Security Investigation into TikTok – Sources," Reuters, 1 Nov. 2019, <https://www.reuters.com/article/us-tiktok-cfius-exclusive/exclusive-u-s-opens-national-security-investigation-into-tiktok-sources-idUSKBN1XB4IL>.
- 116 Marco Silva, "Video App TikTok Fails to Remove Online Predators," BBC News, 5 Apr. 2019, <https://www.bbc.com/news/blogs-trending-47813350>.
- 117 Sam Biddle, Paulo Victor Ribeiro, and Tatiana Dias, "Invisible Censorship: TikTok Told Moderators to Suppress Posts by 'Ugly' People and the Poor to Attract New Users," The Intercept, 15 Mar. 2020, <https://theintercept.com/2020/03/16/tiktok-app-moderators-users-discrimination/>.
- 118 "Advocacy Group Says TikTok Violated FTC Consent Decree and Children's Privacy Rules," Reuters, 13 May 2020, <https://www.yahoo.com/now/advocacy-group-says-tiktok-violated-040831858.html>.
- 119 To "set the record straight," the company explained that it stores all its data in the U.S., not China, and that it has a "dedicated technical team focused on adhering to robust cybersecurity policies, and data and privacy and security practices." The statement said that the company "does not remove content based on sensitivities related to China.... We are not influenced by any foreign government, including the Chinese government...." TikTok, "Statement on TikTok's Content Moderation and Data Security Practices," 24 Oct. 2019, <https://newsroom.tiktok.com/en-us/statement-on-tiktoks-content-moderation-and-data-security-practices>.
- 120 Culliford, "TikTok Reveals First Members of New U.S. Content-moderation Committee."
- 121 TikTok, "TikTok Transparency Center," <https://www.tiktok.com/transparency/en-us/>.
- 122 Sam Shead, "TikTok is Luring Facebook Moderators to Fill New Trust and Safety Hubs," CNBC, 12/Nov. 2020/, <https://www.cnbc.com/2020/11/12/tiktok-luring-facebook-content-moderators.html>.
- 123 Eric Han, "Protecting Against Exploitative Content," TikTok, 5 Mar. 2020, <https://newsroom.tiktok.com/en-us/protecting-against-exploitative-content>.
- 124 Vanessa Pappas, "Strengthening Our Commitment to Transparency," TikTok, 27 July 2022, <https://newsroom.tiktok.com/en-us/strengthening-our-commitment-to-transparency>.
- 125 Integral Ad Science, "TikTok Brand Safety Solution: Make Every Impression Count"; Garrett Sloane, "How TikTok is Trying to Guarantee Brand Safety on its 'For You' Page," *Ad Age*, 22 Nov. 2021, <https://adage.com/article/digital-marketing-ad-tech-news/tiktok-aims-brand-safety-you-page/2382061>.
- 126 TikTok, "TikTok Partners with Zefr on Brand Safety and Brand Suitability Measurement," 24 Jan. 2022, <https://www.tiktok.com/business/en-US/blog/zefr-partnership-brand-safety-suitability-measurement>; "Zefr Partners with TikTok to Provide Brand Safety and Brand Suitability Measurement to Advertisers, Aligned to GARM Industry Standards," PR Newswire, 24 Jan. 2022, <https://www.prnewswire.com/news-releases/zefr-partners-with-tiktok-to-provide-brand-safety-and-brand-suitability-measurement-to-advertisers-aligned-to-garm-industry-standards-301466195.html>.
- 127 Han, "Protecting Against Exploitative Content."
- 128 For example, children are not allowed to exchange messages with others and their profiles are private. "Privacy Policy for Younger Users," TikTok, Jan. 2020, <https://www.tiktok.com/legal/privacy-policy-for-younger-users?lang=en>.

- 129 Tracy Elizabeth, "Our Work to Design an Age-appropriate Experience on TikTok," TikTok, 12 May 2021, <https://newsroom.tiktok.com/en-us/our-work-to-design-an-age-appropriate-experience-on-tiktok>; TikTok, "Safety Center: Guides," <https://www.tiktok.com/safety/en-us/>.
- 130 TikTok, "Controlling What People See on Your Profile," 8 May, 2019, <https://newsroom.tiktok.com/en-us/post-7-controlling-what-people-see-on-your-profile>; Jessica Conditt, "TikTok Rolls Out a Guardian's Guide to Help Parents Understand the Platform," Yahoo! Finance, 11 May 2021, <https://finance.yahoo.com/news/tiktok-safety-guardians-guide-may-2021-180657064.html>; TikTok, "Account Settings," <https://www.tiktok.com/safety/en-nz/account-settings/>.
- 131 Conditt, "TikTok Rolls Out a Guardian's Guide to Help Parents Understand the Platform"; TikTok, "Guardian's Guide," <https://www.tiktok.com/safety/en/guardians-guide/>; TikTok, "Bullying Prevention," <https://www.tiktok.com/safety/en-us/bullying-prevention/>.
- 132 Faith Karimi, "TikTok Disables Late-night Notifications for Teens as Part of New Safety Measures," CNN Business, 12 Aug. 2021, <https://www.cnn.com/2021/08/12/business/tiktok-child-safety-measures/index.html>; Alexandra Evans, "Furthering Our Safety and Privacy Commitments for Teens on TikTok," TikTok, 12 Aug. 2021, <https://newsroom.tiktok.com/en/furthering-our-safety-and-privacy-commitments-for-teens-on-tiktok-us>.
- 133 Daniel Konstantinovic, "TikTok Releases Report on Teen Safety as Platforms Try to Preempt Further Regulator Scrutiny," eMarketer, 19 Nov. 2021, <https://www.emarketer.com/content/tiktok-releases-report-on-teen-safety-platforms-try-preempt-further-regulator-scrutiny>.
- 134 Cormac Keenan, "TikTok Proposes Global Coalition to Protect Against Harmful Content," TikTok, 26 Oct. 2020, <https://newsroom.tiktok.com/en-ie/tiktok-global-coalition-protect-against-harmful-content>.
- 135 INHOPE, "TikTok Supports INHOPE in the Fight Against CSAM," 2 Oct. 2021, <https://www.inhope.org/EN/articles/tiktok-supports-inhope-in-the-fight-against-csam>.
- 136 Cecilia Kang, "A Coalition of State Attorneys General Opens an Investigation into TikTok," *New York Times*, 2 Mar. 2022, <https://www.nytimes.com/2022/03/02/technology/tiktok-states-investigation.html?referringSource=articleShare>.
- 137 "ICO Could Impose Multi-million Pound Fine on TikTok for Failing to Protect Children's Privacy," Information Commissioner's Office, 26 Sept. 2022, <https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/09/ico-could-impose-multi-million-pound-fine-on-tiktok-for-failing-to-protect-children-s-privacy/>.
- 138 Twitch has also expanded its offerings to include cooking, music, and other content. Joseph Yaden, "What is Twitch?" Digital Trends, 25 Oct. 2020, <https://www.digitaltrends.com/gaming/what-is-twitch/>; Devon Delfino, "'What is Twitch?': Here's What You Need to Know About the World's Leading Live-Streaming Platform for Gamers," Business Insider, 11 June 2020, <https://www.businessinsider.com/what-is-twitch>.
- 139 Scott Elchison, "The Difference Between Esports & Gaming," IPG Media Lab, 11 Jul. 2019, <https://medium.com/ipg-media-lab/the-difference-between-esports-gaming-3bcb2d45d42b>.
- 140 Nielsen, "Esports Playbook for Brands 2019," <https://www.nielsen.com/wp-content/uploads/sites/3/2019/05/esports-playbook-for-brands-2019.pdf>; Nielsen Sports, "Riot Games Taps Nielsen for Sponsorship Measurement," 27 June 2019, <https://niensports.com/riot-games-taps-nielsen-for-sponsorship-measurement/>; SuperData, "Arcade," <https://www.superdataresearch.com/arcade>.
- 141 Tarpley Hitt and Emily Shugerman, "Inside the 'Wild West' of Twitch Video-Game Streaming, Where Sexual Abuse Runs Rampant," The Daily Beast, 25 June 2020, <https://www.thedailybeast.com/inside-the-wild-west-of-twitch-video-game-streaming-where-sexual-abuse-runs-rampant>. See also, Matt Perez, "Top-Earning Video Gamers: The Ten Highest-Paid Players Pocketed More Than \$120 Million In 2019," *Forbes*, 29 Jan. 2020, <https://www.forbes.com/sites/mattperez/2020/01/29/top-earning-video-gamers-the-ten-highest-paid-players-pocketed-more-than-120-million-in-2019/#2bccd74f4880>.
- 142 Twitch Advertising, "Audience: Over 2,500,000 are Watching Twitch Right Now," <https://twitchadvertising.tv/audience/>.
- 143 "Mobile Gaming Gains During Lockdown, E-sports Unchanged," WARC, 28 July 2020, <https://www.warc.com/newsandopinion/news/mobile-gaming-gains-during-lockdown-e-sports-unchanged/43903>.
- 144 "What We Know about Marketing via Gaming," WARC Best Practice, Sept, 2020, <https://www.warc.com/content/article/what-we-know-about-marketing-via-gaming/111125> (subscription required); eMarketer, "US Consumer Spending on Video Games, Video Streaming, and Audio Streaming, 2021 & 2022, (billions and % change)," 3 Jan. 2022, <https://chart-na1.emarketer.com/253926/us-consumer-spending-on-video-games-video-streaming-audio-streaming-2021-2022-billions-change> (subscription required); Tom Wijman, "The Games Market and Beyond in 2021: The Year in Numbers," Newzoo, 22 Dec. 2021, <https://newzoo.com/insights/articles/the-games-market-in-2021-the-year-in-numbers-esports-cloud-gaming/>; Dean Takahashi, "SuperData: Games Hit \$120.1 Billion in 2019, with Fortnite Topping \$1.8 Billion," VentureBeat, 2 Jan. 2020, <https://venturebeat.com/2020/01/02/super-data-games-hit-120-1-billion-in-2019-with-fortnite-topping-1-8-billion/>. Nielsen's research, part of its gaming and esports measurement work, notes that for 7-9-year-olds, mobile devices (phones and tablets) are the preferred devices; 10-12-year-olds prefer to play on gaming consoles, such as the PlayStation. SuperData Research, "Preteen Gamers: Attitudes and Habits of the Newest Generation of Gamers," 2019, <https://www.superdataresearch.com/reports/p/preteengamers>; GWI, "The Gaming Playbook: Everything You Need to Know About the Gaming Audience," 2021, [https://www.gwi.com/hubfs/Downloads/The%20Gaming%20Playbook%20-%20GWI.pdf?utm\\_campaign=210413%20The%20gaming%20playbook%20report%20%7Cgwi%20gaming%7C&utm\\_medium=email&\\_hsmsi=121102143&\\_hsenc=p2AN-qtz--KQgoQ1fBDD\\_Wz5Ny2JBjWdaL7F48x8EJqOApTAZ9fIClovzVeRuvPRDeOBe5GW-mDCjCr\\_J9Qmur5n\\_C5jBbQAFaNEQ&utm\\_content=121102143&utm\\_source=hs\\_automation](https://www.gwi.com/hubfs/Downloads/The%20Gaming%20Playbook%20-%20GWI.pdf?utm_campaign=210413%20The%20gaming%20playbook%20report%20%7Cgwi%20gaming%7C&utm_medium=email&_hsmsi=121102143&_hsenc=p2AN-qtz--KQgoQ1fBDD_Wz5Ny2JBjWdaL7F48x8EJqOApTAZ9fIClovzVeRuvPRDeOBe5GW-mDCjCr_J9Qmur5n_C5jBbQAFaNEQ&utm_content=121102143&utm_source=hs_automation).
- 145 Emily Guskin, "Teenagers Are Fueling a Competitive Gaming Tidal Wave," *Washington Post*, 9 Mar. 2018, <https://www.washingtonpost.com/news/sports/wp/2018/03/09/teenagers-are-fueling-an-e-gaming-tidal-wave/>.



- 146 Guskin, "Teenagers Are Fueling a Competitive Gaming Tidal Wave"; Itamar Benedy, "Gaming is the New Social Networking," WARC, 7 June 2020, <https://www.warc.com/newsandopinion/opinion/gaming-is-the-new-social-networking/3711>.
- 147 Adam Epstein, "Twitch is the Undisputed Champion of Video Game Streaming," Quartz, 10 June 2021, <https://qz.com/1966986/twitch-owned-by-amazon-is-the-dominant-force-in-live-streaming/>.
- 148 Twitch Advertising, "Audience: Over 2,500,000 are Watching Twitch Right Now."
- 149 Cecilia D'Anastasio, "Children Stream on Twitch—Where Potential Predators Find Them," *Wired*, 30 July 2020, <https://www.wired.com/story/children-stream-twitch-potential-predators-exploitation>. See also, Bijan Stephen, "The Lockdown Live-streaming Numbers are Out, and They're Huge," *The Verge*, <https://www.theverge.com/2020/5/13/21257227/coronavirus-streamelements-arsenalgg-twitch-youtube-livestream-numbers>; Anthony Ngyuen, "Twitch Releases New, Very Twitch-Specific Content Moderation Policies," *Slate*, 15 Dec. 2019, <https://slate.com/technology/2020/12/twitch-content-moderation-harassment-policy-emotes.html>.
- 150 Amazon, "Diving Deep with Connected Consumers: How to Reach, Engage, and Delight the Streaming Viewers of Today," July 2021, [https://m.media-amazon.com/images/G/01/AdProductsWebsite/docs/Diving\\_Deep\\_with\\_Connected\\_Consumers\\_July\\_2021.pdf](https://m.media-amazon.com/images/G/01/AdProductsWebsite/docs/Diving_Deep_with_Connected_Consumers_July_2021.pdf); Amazon, "Amazon Ads Reaches 55MM+ Monthly Active Viewers of Ad-supported Streaming TV Content," 4 Feb. 2021, <https://advertising.amazon.com/blog/amazon-ott-reaches-55-million-plus-monthly-active-viewers>; Garrett Sloane, "Amazon Puts Display Ads into Twitch Livestreams," *Ad Age*, 26 Oct. 2021, <https://adage.com/article/digital-marketing-ad-tech-news/amazon-puts-display-ads-twitch-livestreams/2376131>; Peter Adams, "Amazon Courts Brand Dollars as Ad-supported OTT Offerings Hit 120M Monthly Viewers," *Marketing Dive*, 4 May 2021, <https://www.marketingdive.com/news/amazon-courts-brand-dollars-as-ad-supported-ott-offerings-hit-120m-monthly/599511/>.
- 151 Arthur Chu, "This Is Not a Game: How SXSW Turned GamerGate Abuse Into a Spectator Sport," *The Daily Beast*, 13 Apr. 2017, <https://www.thedailybeast.com/this-is-not-a-game-how-sxsw-turned-gamer-gate-abuse-into-a-spectator-sport>.
- 152 Jacob Kastrenakes, "Twitch Reckons with Sexual Assault as It Begins Permanently Suspending Streamers," *The Verge*, 25 June 2020, <https://www.theverge.com/2020/6/25/21303185/twitch-sexual-harassment-assault-permanent-bans-streamers>.
- 153 Kastrenakes, "Twitch Reckons with Sexual Assault as It Begins Permanently Suspending Streamers."
- 154 Kastrenakes, "Twitch Reckons with Sexual Assault as It Begins Permanently Suspending Streamers."
- 155 Kastrenakes, "Twitch Reckons with Sexual Assault as It Begins Permanently Suspending Streamers."
- 156 Kastrenakes, "Twitch Reckons with Sexual Assault as It Begins Permanently Suspending Streamers."
- 157 Kastrenakes, "Twitch Reckons with Sexual Assault as It Begins Permanently Suspending Streamers."
- 158 Kastrenakes, "Twitch Reckons with Sexual Assault as It Begins Permanently Suspending Streamers."
- 159 Hitt and Shugerman, "Inside the 'Wild West' of Twitch Video-Game Streaming, Where Sexual Abuse Runs Rampant."
- 160 Hitt and Shugerman, "Inside the 'Wild West' of Twitch Video-Game Streaming, Where Sexual Abuse Runs Rampant."
- 161 Donghee Yvette Wohn, "Volunteer Moderators in Twitch Micro Communities: How They Get Involved, the Roles They Play, and the Emotional Labor They Experience," Association for Computing Machinery, CHI 2019, May 4–9, 2019, Glasgow, Scotland UK, <https://arcticpenguin.files.wordpress.com/2019/02/twitchchi.pdf>.
- 162 Twitch, "Guide to Building a Moderation Team," [https://help.twitch.tv/s/article/guide-to-building-a-moderation-team?language=en\\_US](https://help.twitch.tv/s/article/guide-to-building-a-moderation-team?language=en_US); Twitch, "Moderation on Twitch," <https://www.twitch.tv/creatorcamp/en/setting-up-your-stream/moderation-and-safety/>.
- 163 Twitch, "Hateful Conduct & Harassment," <https://www.twitch.tv/p/en/legal/community-guidelines/harassment/>. See also, Twitch, "Setting Up Moderation for Your Twitch Channel," [https://help.twitch.tv/s/article/setting-up-moderation-for-your-twitch-channel?language=en\\_US](https://help.twitch.tv/s/article/setting-up-moderation-for-your-twitch-channel?language=en_US); Twitch, "How to Use AutoMod," [https://help.twitch.tv/s/article/how-to-use-automod?language=en\\_US](https://help.twitch.tv/s/article/how-to-use-automod?language=en_US).
- 164 Twitch, "Introducing the Twitch Safety Advisory Council."
- 165 Todd Spangler, "Twitch Hires Xbox Exec Angela Hession as Head of Trust and Safety," *Variety*, 10 Nov. 2020, <https://variety.com/2020/digital/news/twitch-xbox-angela-hession-trust-safety-1234827064/>.
- 166 Twitch, "Introducing Our New Hateful Conduct & Harassment Policy," 9 Dec. 2020, [https://blog.twitch.tv/en/2020/12/09/introducing-our-new-hateful-conduct-harassment-policy/?utm\\_referrer=https://safety.twitch.tv/](https://blog.twitch.tv/en/2020/12/09/introducing-our-new-hateful-conduct-harassment-policy/?utm_referrer=https://safety.twitch.tv/)
- 167 Twitch, "Our Work to Combat Online Grooming," [https://safety.twitch.tv/s/article/Our-Work-to-Combat-Online-Grooming?language=en\\_US](https://safety.twitch.tv/s/article/Our-Work-to-Combat-Online-Grooming?language=en_US).
- 168 Bijan Stephen, "Twitch May Be Adding a 'Brand Safety Score' for Streamers," *The Verge*, 9 Mar. 2021, <https://www.theverge.com/2021/3/9/22321311/twitch-api-brand-safety-score-advertising-uber-lyft>.
- 169 Amazon, "Brand Safety Developer Guide," <https://advertising.amazon.com/API/docs/en-us/sponsored-display/brand-safety>; Daylam 'tayari' Tayari, "What is Twitch's Brand Safety Score and What Does it Mean for Streamers?" *Twitter*, 9 Mar. 2021, [https://twitter.com/tayariCS/status/1369230902409097218?ref\\_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1369230902409097218%7Ctwgr%5E%7Ctwcon%5Es1\\_&ref\\_url=https%3A%2F%2F-dotesports.com%2Fstreaming%2Fnews%2Fwhat-is-twitchs-brand-safety-score-and-what-does-it-mean-for-streamers](https://twitter.com/tayariCS/status/1369230902409097218?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1369230902409097218%7Ctwgr%5E%7Ctwcon%5Es1_&ref_url=https%3A%2F%2F-dotesports.com%2Fstreaming%2Fnews%2Fwhat-is-twitchs-brand-safety-score-and-what-does-it-mean-for-streamers); Amazon, "DoubleVerify's Brand Suitability Tiers on Third-party Supply," 19 July 2021, <https://advertising.amazon.com/resources/whats-new/doubleverify-brand-suitability-tiers>.

- 170 As the company explains, “We provide both creators and their mods with a powerful suite of tools such as AutoMod, Chat Modes, and Mod View to make their roles as easy and intuitive as possible. These tools provide the ability to automatically filter chat, allow creators and mods to see (and delete) questionable chat messages before they are displayed on the channel, give users ‘time outs’ (lock them out of chat for a period of time) or permanently block them from the channel.” Twitch, “Transparency Report,” <https://www.twitch.tv/p/en/legal/transparency-report/>. See also, Twitch, “How to Use AutoMod”; Twitch, “Mod View,” <https://help.twitch.tv/s/article/mod-view>. For the 2021 report, see Amazon, “Our Efforts to Combat Child Sexual Abuse Material in 2021,” 16 Mar. 2022, <https://www.aboutamazon.com/news/policy-news-views/our-efforts-to-combat-child-sexual-abuse-material-in-2021>.
- 171 Twitch, “Transparency Report.”
- 172 Twitch, “Transparency Report.”
- 173 National Center on Sexual Exploitation, “Amazon’s Twitch Rife with Sexual Harassment, Predatory Grooming, Child Sexual Abuse,” 2 Mar. 2021, <https://endsexualexploitation.org/articles/amazons-twitch-rife-with-sexual-harassment-predatory-grooming-child-sexual-abuse/>.
- 174 Cecilia D’Anastasio, “Child Predators Use Twitch to Systematically Track Kids Livestreaming,” Bloomberg, 21 Sept. 2022, <https://www.bloomberg.com/graphics/2022-twitch-problem-with-child-predators/?sref=QDmhoVl8#headline-start>; Lily Hay Newman, “Security News This Week: Child Predators Mine Twitch to Prey on Kids,” *Wired*, 24 Sept. 2022, <https://www.wired.com/story/child-predators-twitch-russia-internet-regulator-security-roundup/?redirectURL=/story/child-predators-twitch-russia-internet-regulator-security-roundup/> (subscription required).
- 175 “Twitch is Hiring a Data Scientist, Machine Learning - Community Health,” StartUp.Jobs, <https://startup.jobs/data-scientist-machine-learning-community-health-twitch-2842569>.
- 176 “What We Know about Marketing via Gaming.”
- 177 Rae Paoletta, “How Advertisers Are Using Twitch to Reach People Who Hate Ads,” AdExchanger, 23 Oct. 2018, <https://www.adexchanger.com/ad-exchange-news/how-advertisers-are-using-twitch-to-reach-people-who-hate-ads/>.
- 178 See, for example, Chester, Montgomery, and Kopp, “‘Big Food’ and ‘Big Data’ Online Platforms Fueling Youth Obesity Crisis as Coronavirus Pandemic Rages.”
- 179 “Antitrust: Commission Opens Investigation into Possible Anti-competitive Conduct of Amazon,” 17 July 2019, [https://ec.europa.eu/commission/presscorner/detail/pl/ip\\_19\\_4291](https://ec.europa.eu/commission/presscorner/detail/pl/ip_19_4291); Cecilia Kang and David McCabe, “A House Panel Calls for a Criminal Investigation into Amazon,” *New York Times*, 9 Mar. 2022, <https://www.nytimes.com/2022/03/09/business/amazon-antitrust-investigation.html>.
- 180 Twitch, “Transparency Report.”
- 181 Elise Stegman, “Content Moderation on YouTube and its Effects on Users,” DiggIt Magazine, 20 Sept. 2021, <https://www.diggitmagine.com/articles/content-moderation-youtube>.
- 182 Mara Singer, “Content vs. Devices: What Reigns Supreme?” Kidscreen, 5 Mar. 2019, <http://kidscreen.com/2019/03/05/from-devices-to-content-providers-youtube-reigns-with-kids/>.
- 183 Louise Matsakis, “YouTube Doesn’t Know Where its Own Line Is,” *Wired*, 2 Mar. 2018, <https://www.wired.com/story/youtube-content-moderation-inconsistent/>.
- 184 Elise Stegman, “Content Moderation on YouTube and its Effects on Users.”
- 185 Todd Spangler, “YouTube Faces Advertiser Boycott Over Videos With Kids That Attracted Sexual Predators,” *Variety*, 25 Nov. 2017, <https://variety.com/2017/digital/news/youtube-ad-boycott-pedophile-sexual-children-videos-1202622790/>.
- 186 Max Fisher and Amanda Taub, “On YouTube’s Digital Playground, an Open Gate for Pedophiles,” *New York Times*, 3 June 2019, <https://www.nytimes.com/2019/06/03/world/americas/youtube-pedophiles.html>; K.G. Orphanides, “On YouTube, a Network of Paedophiles is Hiding in Plain Sight,” *Wired*, 20 Feb. 2019, <https://www.wired.co.uk/article/youtube-pedophile-videos-advertising>.
- 187 Fisher and Taub, “On YouTube’s Digital Playground, an Open Gate for Pedophiles.”
- 188 YouTube, “Introducing the Newest Member of Our Family, the YouTube Kids App—Available on Google Play and the App Store,” *YouTube Official Blog*, 23 Feb. 2015, <https://youtube.googleblog.com/2015/02/youtube-kids.html>.
- 189 Jeff Chester, “Advocates Say Google’s YouTube Violates Federal Children’s Privacy Law,” Center for Digital Democracy, 9 Apr. 2018, <https://www.democraticmedia.org/article/advocates-say-googles-youtube-violates-federal-childrens-privacy-law>.
- 190 Kim Hart and Sara Fischer, “The Wild West of Children’s Entertainment,” *Axios*, 6 Aug. 2018, <https://www.axios.com/wild-west-childrens-entertainment-33433183-8fcf-423a-bba2-e3f93d4d38ef.html>.
- 191 Chester, “Advocates Say Google’s YouTube Violates Federal Children’s Privacy Law”; Federal Trade Commission, “Google and YouTube Will Pay Record \$170 Million for Alleged Violations of Children’s Privacy Law,” 4 Sept. 2019, <https://www.ftc.gov/news-events/press-releases/2019/09/google-youtube-will-pay-record-170-million-alleged-violations>; “YouTube Prohibits All Food and Beverage Ads Next to Kids Content as of Today,” World Federation of Advertisers, 10 Jan. 2020, <https://wfanet.org/knowledge/item/2020/01/10/YouTube-prohibits-all-food-and-beverage-ads-next-to-kids-content-as-of-today>; “Senator Markey Commends New Advertising Policies to Protect Children on YouTube,” 18 Dec. 2019, <https://www.markey.senate.gov/news/press-releases/senator-markey-commends-new-advertising-policies-to-protect-children-on-youtube>.

- 192 Kristin Cohen, "YouTube Channel Owners: Is Your Content Directed to Children?" *FTC Business Blog*, 22 Nov. 2019, <https://www.ftc.gov/business-guidance/blog/2019/11/youtube-channel-owners-your-content-directed-children>; YouTube, "Better Protecting Kids' Privacy on YouTube," *YouTube Official Blog*, 6 Jan. 2020, <https://blog.youtube/news-and-events/better-protecting-kids-privacy-on-youtube/>; Susan Wojcicki, "An Update on Kids and Data Protection on YouTube," *YouTube Official Blog*, 4 Sept. 2019, <https://blog.youtube/news-and-events/an-update-on-kids/>.
- 193 Todd Spangler, "Google's YouTube to Pay \$170 Million to Settle FTC Charge It Collected Kids' Data Illegally," *Variety*, 4 Sept. 2019, <https://variety.com/2019/digital/news/youtube-ftc-fine-170-million-coppa-kids-data-1203323004/>. See also, "Child-friendly Changes Coming to Google Search and YouTube," *Engineering and Technology*, 10 Aug. 2021, <https://eandt.theiet.org/content/articles/2021/08/child-friendly-changes-coming-to-google-search-and-youtube/>; 5Rights Foundation, "Google Announcement Shows Impact of Children's Code," 10 Aug. 2021, <https://5rightsfoundation.com/in-action/google-announcement-shows-impact-of-childrens-code.html>.
- 194 James Beser, "New Safety and Digital Wellbeing Options for Younger People on YouTube and YouTube Kids," *YouTube Official Blog*, 10 Aug. 2021, <https://blog.youtube/news-and-events/new-safety-and-digital-wellbeing-options-younger-people-youtube-and-youtube-kids/>.
- 195 Casey Newton, "The Terror Queue," *The Verge*, 16 Dec. 2019, <https://www.theverge.com/2019/12/16/21021005/google-youtube-moderators-ptsd-accenture-violent-disturbing-content-interviews-video>.
- 196 Will Oremus, "'YouTube Magic Dust': How America's Second-largest Social Platform Ducks Controversies," *Washington Post*, 25 Aug. 2021, <https://www.washingtonpost.com/technology/2021/08/25/youtube-content-moderation-strategy/>.
- 197 Google, "YouTube Community Guidelines Enforcement," <https://transparencyreport.google.com/youtube-policy/removals?hl=en>; YouTube, "The Four Rs of Responsibility, Part 2: Raising Authoritative Content and Reducing Borderline Content and Harmful Misinformation," *YouTube Official Blog*, 3 Dec. 2019, <https://blog.youtube/inside-youtube/building-greater-transparency-and-accountability/>; Rebecca Heilweil, "YouTube Says it's Better at Removing Videos that Violate its Rules, but Those Rules Are in Flux," *Vox*, 6 Apr. 2021, <https://www.vox.com/recode/2021/4/6/22368809/youtube-violative-view-rate-content-moderation-guidelines-spam-hate-speech>. See also, Kim Lyons, "YouTube Took Down More Videos Than Ever Last Quarter as it Relied More on Non-human Moderators," *The Verge*, 25 Aug. 2020, <https://www.theverge.com/2020/8/25/21401435/youtube-videos-moderators-filters-human-appeals>; Google, "Spam, Deceptive Practices, & Scams Policies," YouTube Help, <https://support.google.com/youtube/answer/2801973?hl=en#:~:text=YouTube%20doesn't%20allow%20spam,violates%20this%20policy%2C%20report%20it>; Google, "COVID-19 Medical Misinformation Policy," YouTube Help, [https://support.google.com/youtube/answer/9891785?hl=en&ref\\_topic=9282436](https://support.google.com/youtube/answer/9891785?hl=en&ref_topic=9282436).
- 198 Heilweil, "YouTube Says it's Better at Removing Videos that Violate its Rules, but Those Rules Are in Flux."
- 199 YouTube also has a separate metric called the Advertiser Safety Rate to determine how safe its platform is for marketers by identifying content deemed not suitable for advertising. A YouTube representative said the company had 20,000 people working on brand-safety-related issues in 2021. Pixability, "Understanding Brand Suitability's Relationship with Connected TV Advertising | Webinar," YouTube, 22 Nov. 2021, [https://www.youtube.com/watch?v=dhr5\\_SqOIl0](https://www.youtube.com/watch?v=dhr5_SqOIl0); YouTube, "The Four Rs of Responsibility, Part 2."
- 200 Google, "Google Transparency Report," <https://transparencyreport.google.com/?hl=en>; Google, "Advertiser-friendly Content Guidelines," YouTube Help, <https://support.google.com/youtube/answer/6162278>; Google, "YouTube's Community Guidelines," YouTube Help, <https://support.google.com/youtube/answer/9288567>; Google, "Brand Safety Targeting," Display & Video 360 Help, <https://support.google.com/displayvideo/answer/3032915?hl=en>; Google, "We've Got Your Back," Google Ad Manager, <https://admanager.google.com/home/capabilities/brand-safety/>; Google, "Ad Manager and Ad Exchange Program Policies: Tag a Site or Ad Request for Child-directed Treatment," Google Ad Manager Help, <https://support.google.com/admanager/answer/4442399?hl=en>; Google, "Ad Manager and Ad Exchange Program Policies: Google Ad Manager Partner Guidelines," Google Ad Manager Help, <https://support.google.com/admanager/answer/9059370?hl=en>; Google, "Child Sexual Abuse and Exploitation," Publisher Policies Help, <https://support.google.com/publisherpolicies/answer/10436971>.
- 201 Google, "Sexual Content," Advertising Policies Help, <https://support.google.com/adspolicy/answer/6023699>; Google, "Sensitive Categories in Display & Video 360 Brand Safety Targeting," Display & Video 360 Help, <https://support.google.com/displayvideo/answer/6327207>. Note also the process Google uses with Integral Ad Science: Google, "About Integral Ad Science's Media Quality Verification," Display & Video 360 Help, [https://support.google.com/displayvideo/answer/3297897?hl=en&ref\\_topic=4390839](https://support.google.com/displayvideo/answer/3297897?hl=en&ref_topic=4390839).
- 202 Google, "Ads & Made for Kids Content," [https://support.google.com/adspolicy/answer/9683742?hl=en&ref\\_topic=1626336](https://support.google.com/adspolicy/answer/9683742?hl=en&ref_topic=1626336); YouTube, "Policies & Guidelines," YouTube Creators, [https://www.youtube.com/creators/how-things-work/policies-guidelines/?utm\\_source=paid-search&utm\\_medium=gyt&utm\\_id=ytgen&utm\\_content=ytcpg&gclid=Cj0KCQjwma6TBhDIARIsAOKuANwmMVnmsUePxDghZG\\_BiseJEzI\\_SaugJhoTNg6wT82a7SzdMvBNXgaArcxEALw\\_wcB](https://www.youtube.com/creators/how-things-work/policies-guidelines/?utm_source=paid-search&utm_medium=gyt&utm_id=ytgen&utm_content=ytcpg&gclid=Cj0KCQjwma6TBhDIARIsAOKuANwmMVnmsUePxDghZG_BiseJEzI_SaugJhoTNg6wT82a7SzdMvBNXgaArcxEALw_wcB).
- 203 See Katherine C. Montgomery, "Youth and Surveillance in the Facebook Era: Policy Interventions and Social Implications," *Telecommunications Policy* 39, no. 3 (2015), pp. 771-786, <https://www.sciencedirect.com/science/article/abs/pii/S0308596114001955>. See also, Tijana Milosevic, *Protecting Children Online?: Cyberbullying Policies of Social Media Companies* (Cambridge, MA: The MIT Press, 2018).

- 204 See, for example, Rohith Jyothish, "How India's New Internet Regulations Will Change Social Media, Online News and Video Streaming," *Global Voices*, 16 Mar. 2021, <https://globalvoices.org/2021/03/16/how-indias-new-internet-regulations-will-change-social-media-online-news-and-video-streaming/>; Chris Buckley, "China Tightens Limits for Young Online Gamers and Bans School Night Play," *New York Times*, 30 Aug. 2021, <https://www.nytimes.com/2021/08/30/business/media/china-online-games.html>.
- 205 YWCA, "YWCA is on a Mission: Child Sexual Abuse Facts," Sept. 2017, <https://www.ywca.org/wp-content/uploads/WWW-CSA-Fact-Sheet-Final.pdf>; Yeong Yeong Lim, Suzaily Wahab, Jaya Kumar, Fauziah Ibrahim, and Mohammad Rahim Kamaluddin, "Typologies and Psychological Profiles of Child Sexual Abusers: An Extensive Review," *Children* 8, no. 5 (May 2021), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8146192/>.
- 206 Jane Chung, "Big Tech, Big Cash: Washington's New Power Players," *Public Citizen*, 24 Mar. 2021, <https://www.citizen.org/article/big-tech-lobbying-update/>.
- 207 The White House, "State of the Union Address," <https://www.whitehouse.gov/state-of-the-union-2022/>.
- 208 The White House, "Remarks of President Joe Biden—State of the Union Address As Prepared for Delivery," 1 Mar. 2022, <https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/03/01/remarks-of-president-joe-biden-state-of-the-union-address-as-delivered/>; The White House, "FACT SHEET: President Biden to Announce Strategy to Address Our National Mental Health Crisis, As Part of Unity Agenda in his First State of the Union," 1 Mar. 2022, <https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/01/fact-sheet-president-biden-to-announce-strategy-to-address-our-national-mental-health-crisis-as-part-of-unity-agenda-in-his-first-state-of-the-union/>.
- 209 The White House, "FACT SHEET: President Biden to Announce Strategy."
- 210 The White House, "FACT SHEET: President Biden to Announce Strategy."
- 211 "S.1628 - Children and Teens' Online Privacy Protection Act," 117th Congress (2021-2022), <https://www.congress.gov/bill/117th-congress/senate-bill/1628/text>.
- 212 "Blumenthal & Blackburn Introduce Comprehensive Kids' Online Safety Legislation," 16 Feb. 2022, <https://www.blumenthal.senate.gov/newsroom/press/release/blumenthal-and-blackburn-introduce-comprehensive-kids-online-safety-legislation>.
- 213 "Senators Markey and Blumenthal, Rep. Castor Reintroduce Legislation to Protect Children and Teens from Online Manipulation and Harm," 30 Sept. 2021, <https://www.markey.senate.gov/news/press-releases/senators-markey-and-blumenthal-rep-castor-reintroduce-legislation-to-protect-children-and-teens-from-online-manipulation-and-harm/>; "Rep. Castor Reintroduces Landmark Kids PRIVCY Act to Strengthen COPPA, Keep Children Safe Online," 29 July, 2021, <https://castor.house.gov/news/documentsingle.aspx?DocumentID=403677>; "Senators Markey and Cassidy Propose Bipartisan Bill to Update Children's Online Privacy Rules," 24 June 2021, <https://www.markey.senate.gov/news/press-releases/senators-markey-and-cassidy-propose-bipartisan-bill-to-update-childrens-online-privacy-rules>; Common Sense, "The Kids Act: Kids Internet Design and Safety Act," [https://www.common sense media.org/sites/default/files/featured-content/files/kids\\_act\\_fact\\_sheet\\_2021\\_1.pdf](https://www.common sense media.org/sites/default/files/featured-content/files/kids_act_fact_sheet_2021_1.pdf).
- 214 "Senators Markey, Sasse, Blunt, Schatz, Collins, Bennet and Reps. Raskin, Gonzalez, Trahan Reintroduce Bipartisan, Bicameral Legislation to Study Impact of Technology and Media on Children," 14 July 2021, <https://www.markey.senate.gov/news/press-releases/senators-markey-sasse-blunt-schatz-collins-bennet-and-reps-raskin-gonzalez-trahan-reintroduce-bipartisan-bicameral-legislation-to-study-impact-of-technology-and-media-on-children>.
- 215 "AB-2273 The California Age-Appropriate Design Code Act. (2021-2022)," [https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill\\_id=202120220AB2273](https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202120220AB2273).
- 216 "Portman, Colleagues Introduce EARN IT Act to Combat Online Child Sexual Exploitation," 1 Feb. 2022, <https://www.portman.senate.gov/newsroom/press-releases/portman-colleagues-introduce-earn-it-act-combat-online-child-sexual>; Joe Mullin, "It's Back: Senators Want EARN IT Bill to Scan All Online Messages," *Electronic Frontier Foundation*, 3 Feb. 2022, <https://www.eff.org/deeplinks/2022/02/its-back-senators-want-earn-it-bill-scan-all-online-messages>.
- 217 Issie Lapowsky and Emily Birnbaum, "Mark Warner is Ready to Fight for Section 230 Reform," *Protocol*, 22 Mar. 2021, <https://www.protocol.com/policy/mark-warner-section-230>.
- 218 "Wyden, Gillibrand, Brown, Hirono and Eshoo Reintroduce Invest in Child Safety Act to Protect Children from Online Exploitation," 4 Feb. 2021, <https://www.wyden.senate.gov/news/press-releases/wyden-gillibrand-brown-hirono-and-eshoo-reintroduce-invest-in-child-safety-act-to-protect-children-from-online-exploitation>.
- 219 Cat Zakrzewski, "A Bill Aiming to Protect Children Online Reignites a Battle Over Privacy and Free Speech," *Washington Post*, 10 Feb. 2022, <https://www.washingtonpost.com/technology/2022/02/10/senators-earn-it-privacy-children-safety/>; Mullin, "It's Back: Senators Want EARN IT Bill to Scan All Online Messages."
- 220 Jon Porter, "Apple Scrubs Controversial CSAM Detection Feature from Webpage but Says Plans Haven't Changed," *The Verge*, 15 Dec. 2021, <https://www.theverge.com/2021/12/15/22837631/apple-csam-detection-child-safety-feature-webpage-removal-delay>.
- 221 See Valerie Verdoodt, "The Role of Children's Rights in Regulating Digital Advertising," *The International Journal of Children's Rights* 27, no. 3, (2019); 455-481, doi: <https://doi.org/10.1163/15718182-02703002>.



- 222 EU Parliament, “Digital Services Act: Agreement for a Transparent and Safe Online Environment,” 23 Apr. 2022, <https://www.europarl.europa.eu/news/de/press-room/20220412IPR27111/digital-services-act-agreement-for-a-transparent-and-safe-online-environment>; Ryan Browne, “EU Agrees on Landmark Law Aimed at Forcing Big Tech Firms to Tackle Illegal Content,” CNBC, 22 Apr. 2022, <https://www.cnbc.com/2022/04/22/digital-services-act-eu-agrees-new-rules-for-tackling-illegal-content.html>.
- 223 See, for example, leading global marketing agency WPP on better protecting youth: “WPP and Epic Games Partner to Accelerate Innovation for Clients in the Metaverse,” Business Wire, 5 May 2022, <https://www.businesswire.com/news/home/20220504005943/en/WPP-and-Epic-Games-Partner-to-Accelerate-Innovation-for-Clients-in-the-Metaverse>.
- 224 This is a principle that was built into the 1998 Children’s Online Privacy Protection Act (COPPA), which has enabled the Federal Trade Commission to update the regulations emanating from that statute as technologies and online practices have changed over the years. Federal Trade Commission, “16 CFR Part 312, RIN 3084–AB20, Children’s Online Privacy Protection Rule,” *Federal Register* 78, n. 12, (17 Jan. 2013): 3972–4014, <https://www.ftc.gov/system/files/2012-31341.pdf>.
- 225 YouthPower, “Positive Youth Development (PYD) Framework,” <https://www.youthpower.org/positive-youth-development-pyd-framework>.
- 226 UNICEF, “Children’s Rights and the Internet,” <https://www.unicef.org/csr/childrensrightsandinternet.htm>. UNICEF also cites the original United Nations Human Rights document, “Convention on the Rights of the Child,” 20 Nov. 1989, <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child>.