European Journal of Criminology

Article

European Journal of Criminology
1–24
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14773708221131659
journals.sagepub.com/home/euc
SAGE

Criminogenic and harm-enabling features of social media platforms: The case of sharenting practices

Anita Lavorgna
Department of Sociology, Social Policy and Criminology, University of Southampton, UK; Department of Political and Social Sciences, University of Bologna, Bologna, Italy

Morena Tartari
Department of Sociology, Social Policy and Criminology, University of Southampton, UK; Department of Philosophy, Sociology, Education and Applied Psychology, University of Padova, Padova, Italy

Pamela Ugwudike
Department of Sociology, Social Policy and Criminology, University of Southampton, Southampton, UK

Abstract

Sharenting – that is, the sharing of identifying and sensitive information of minors, who are often overexposed online by parents or guardians – has, at times, criminogenic potential, as the information shared can enable both heinous crimes and other types of harmful conduct. Whilst most research on sharenting has focused on the sharenters and their agency, there is a gap in addressing whether and to what extent social media platforms display criminogenic or other harm-enabling features that can render sharenting risky for affected minors. By relying on an adapted crime proofing of legislation approach, our contribution analyses the self-regulations (in the form of corporate documents and forms of self-organisation) of five major social media platforms and identifies several risks and vulnerabilities to harmful sharenting practices embedded in the platforms' policies. In doing so, the study demonstrates how criminological imagination can effectively contribute to the multidisciplinary debates on digital ecosystems and their regulation, paving the way for a reduction of criminogenic and harmful opportunities online.

Corresponding author:
Anita Lavorgna, Department of Sociology, Social Policy & Criminology, University of Southampton, Murray Building (Bldg 58), SO17 1BJ UK.
Email: anita.lavorgna@unibo.it

Keywords
Sharenting, digital harms, crime proofing of legislation, self-regulation, social media platforms

Introduction

As digital technologies continue to permeate most aspects of social life from work and social networking to criminal justice, they are creating 'digital paradoxes' by enabling positive social changes and improvements to the quality of life, whilst generating new threats to safety and wellbeing (Fussey and Roth, 2020). These new techno-socialities (Escobar, 1994) are particularly interesting from a criminological perspective, beyond the mainstream cybercrime literature, as individuals interacting in non-criminal virtual spaces can give rise to criminal, deviant or otherwise harmful activities and behaviours (Powell et al., 2018; Lavorgna, 2021). In this context, mainstream social media platforms play a key role, providing highly populated virtual spaces. These platforms are best understood as sociotechnical assemblages and complex institutions (Gillespie, 2017), and are often conceptualised as composite human actors (users and, depending on the platform, moderators) and algorithm-driven nonhuman entities (automated tools and filters) embedded in their users' general communicative practices (in line with Prochazka, 2019).

In this contribution which originates from the ESRC-funded project *ProTechThem - Building Awareness for Safer and Technology-Savvy Sharenting*, we hypothesise that virtual places have regulatory characteristics and loopholes that can enable harmful practices amongst users. We demonstrate this using the empirical example of harmful sharenting practices in online social media spaces. Sharenting is the potentially harmful sharing of identifying and sensitive information of minors, who are often overexposed online by parents or guardians.

Whilst most research on sharenting has focused on the sharenters and their agency, the characteristics of the social media platforms they inhabit, which are at the basis of their potentially harmful sharing activities, have been so far overlooked. To address this gap, we draw on insights from the ESRC project which analysed the self-regulatory policies and practices of five mainstream social media platforms. The aim was to examine whether and to what extent the platforms display problematic features, and whether there are regulatory loopholes that can render sharenting risky for affected children by enabling criminogenic and other harm-enabling opportunities. To achieve our objective, we utilised an adapted crime proofing of legislation approach which is a conceptual framework for identifying criminogenic opportunities in regulations and their implementations (see Savona, 2017). This allowed us to examine whether criminogenic or otherwise harm-enabling features are embedded in the policies. Of course, what we are addressing in our work is a complex sociotechnical phenomenon, based on the interplay between policies, platform design and developments, and social norms and attitudes. We do not aim to provide a comprehensive understanding of the issue, but rather to shed some light on one of the factors accounting for important (criminal and non-criminal) harms and risks that social media platforms are currently failing to address – that is, their self-regulatory practices. As such, in this contribution we are focusing on the orientation of social media platforms towards the legal risks relating to sharenting, rather than, for instance, on other relevant factors such as their design. In the next section, we offer a brief overview of the nature and limitations of self-regulation (via content moderation) as a core governance mechanism of social media platforms.

Framing the ecosystem: Self-regulation and moderation in social-media platforms

A key dilemma of contemporary digital policy relates to the issue of how best regulate online media, particularly social media platforms, as they are fundamental digital intermediaries in contemporary societies (Milosavljević and Micova, 2016; Cusumano et al., 2021). At the moment, self-regulatory mechanisms (and especially corporate documents and forms of self-organisation, consider for instance the terms of service or content policies created by the social media companies themselves) play a core role. After all, self-regulatory initiatives are not novel developments, as they are often used by industries to improve their public reputation (also relatively to their competition), and to avoid more costly regulation (Abbott and Snidal, 2009; Milosevic, 2017; Gorwa, 2019).

Of course, self-regulation alone is not enough; appropriate regulatory procedures are also required (consider, for instance, Zancova and Dimitrov, 2020; Tyler et al., 2021). More generally, there have been calls for increased accountability of social media platforms (for instance, see Leerssen, 2015) and for better regulations and controls backed by a strong sanction regime against misbehaving platforms (Wise, 2019). Concerns over the data practices of social media platforms and privacy violations, for instance, have led to harsh criticisms, with claims that they should not be permitted to behave like ‘digital gangsters’ in the online world, considering themselves to be ahead of and beyond the law (DCMS, 2019: 42). Beside self-regulation, in the European region, several multi-actor informal arrangements governing online content on social media platforms are in place. They are likely to play an increasingly important role in the coming years as a vital part of the corporate regulatory toolbox (Gorwa, 2019), which can help avoid issues of techlash (Douek, 2019). In recent years, social media platforms seem to be entering a new phase of responsibility towards the public, but solutions to commonly cited problems such as privacy violations and poor content moderation have not been offered. Thus, current self-regulatory strategies seem to constitute nothing more than a smokescreen or *digital washing.* Nevertheless, they remain the primary form of regulation and are, as such, pivotal to understanding the criminogenic features of the platforms and how they are addressed. Self-regulation also mostly defines the relationship between social media platforms and their users – that is, the digital ecosystem of our interest.

As digital intermediaries, social media platforms are constantly muddling through balancing acts between users’ ability to post freely (promoting individual participation and fostering discourse democracy) and preventing harms that may arise from posted content, whilst operating in a context whereby the platform is commodifying users’ content and data to make a profit (Johnson, 2017). Self-regulatory practices can favour this balance, creating a trusted partnership between users and social media companies (Schneble et al., 2021). But they still fall short, operating in a ‘logic of opacity’ (Roberts, 2018). It has been claimed, for instance, that terms and conditions, and rule of conducts, remain obscure to users, as regulations are written in long and complex language (Schneble et al., 2021).

In the context of self-regulatory practices, content moderation can be defined as the ‘governance mechanisms that structure participation in a community to facilitate cooperation and prevent abuse' (Grimmelmann, 2015: 6). Moderation can take different forms: it can be automatised or manual, and it can range from centralised corporate moderation to user-driven moderation models (Seering, 2020). Moderation can have both a punitive and an educational role (West, 2018). Either way, moderation systems have a role in shaping affective relationship between users and platforms, and for users to assert their agency, for instance by seeking redress (West, 2018). As such, they are a first point of intervention for preventing and countering several criminal, deviant, or otherwise harmful activities online, in the effort to create, nurture or maintain a better social media platform ecosystem. Following Seering's (2020) distinction among 'the platforms and policies perspective' and 'the communities perspective' in moderation research, acknowledging that previous research has already stressed how platforms generally allow users significant leeway to self-moderate (for a review, see Seering, 2020), and fully recognising the importance of intra-group moderation, this study departs by looking at platforms and policies as – we claim – these provide the overarching frame within which also users' intra-group moderation occurs.

Over the years, automated ways to moderate social media have been developed by social media companies, mostly through artificial intelligence (machine learning) tools based on natural language processing and sentiment analysis, with the intent of removing 'bad' content more effectively and quickly (see, for instance, Gorwa et al., 2020; Lim et al., 2020). These algorithmic moderation systems are increasingly used for user-generated content moderation at scale by all the major social media platforms, fuelled by growing public expectations for increased platform responsibility (for an excellent recent overview of algorithmic moderation, see Gorwa et al., 2020). Unfortunately, these systems are still opaque, unaccountable, and scarcely effective in complex socio-technical systems, with the risk of further complicating and exacerbating issues of (lack of) fairness and justice (Gorwa et al., 2020) as it is extremely difficult for automated tools to make contextual decisions on composite, multifactorial concepts (Li and Williams, 2018). Additionally, intervening with an automated mechanism against something in violation of a platform's standards does not mitigate concerns on how certain standards are created, as at the core of these concerns are issues of transparency, fairness and depoliticisation ('where do we draw the line between what is acceptable and what is not?') that are socio-political and ethical, rather than technical, in nature (as discussed in detail in Gorwa et al., 2020). Yet, as stressed by Gorwa and colleagues (2020), these systems are here to stay, as they are now often mandated by both legislation and informal platform regulation.

In commercial content moderation, the involvement of more humane, trained and diverse moderators in the process is often seen as a practical solution; when manual moderation occurs, however, this is nonetheless done in less-than-ideal conditions, often by (outsourced) freelancers with poor working conditions, and exposed to extreme amounts of toxic content (Milosevic, 2017; Gillespie, 2018). Furthermore, also in those cases, moderation processes are generally still obscure, even if data disclosures are emerging (Gillespie, 2018; Keller and Leerssen, 2019). Opacity, in a way, seems to be a distinguishing feature of platform design, as content moderation practices are entangled in a nebulous web of rules and procedures. In this context, the key criterion in commercial moderation seems to be the potentially revenue-generation value of a certain content, rather than its meaning and intent; as such, platforms become an instrument for the reification and consolidation of pre-existing power structures (Roberts, 2018; Santos Rutschman, 2021).

From this brief excursus, it is evident that analysing self-regulatory mechanisms is necessary for assessing whether and to what extent social media companies are addressing the criminogenic or otherwise harming potential of their platforms. As already noted, we use harmful sharenting practices to explore this issue empirically, with a focus on whether the self-regulatory strategies instituted by social media platforms encompass potentially problematic features that can trigger or exacerbate sharenting risks and harms.

The next section will provide an overview of sharenting practices, risks, and harms. The section will clarify why these matter from a criminological perspective and present necessary multidisciplinary information.

Sharenting and its ecosystem

Sharenting can be defined as 'making public by parents a lot of detailed information about their children in the form of photos, videos and posts through social media, which violate children's privacy' (Brosch, 2018: 78). Especially over the last decade, sharenting has received scholarly attention by several disciplines, including law (e.g., Steinberg, 2017; Hancock, 2021), media, communication and cultural studies (e.g., Chalken and Anderson, 2017; Choi and Lewallen, 2018; Archer, 2019; Ranzini et al., 2020; Barnes and Potter, 2021); computer science (e.g., Ammari et al., 2015); educational sciences (e.g., Cino and Damozzi, 2017; Di Bari, 2017; Brosch, 2018); and psychology (e.g., Lazard et al., 2019). So far, however, this common social practice has been almost ignored by criminological scholarship (Lavorgna et al., 2022), which is surprising as we consider that, beyond risks posed by negative psychological repercussions in ignoring children's desire to having (or not) an online identity (Steinberg, 2017) or due to the perpetuation of gender and racial stereotypes (Choi and Lewallen, 2018), there are concerns regarding the potential for financial exploitation (Archer, 2019; Barassi, 2019), grooming and child abuse, cyber hate and identity crimes (e.g., Bezáková et al., 2021; Wachs et al., 2021; Williams-Ceci et al., 2021). Indeed, recent research has shown that, despite the potential under-emerging and underreporting of cases where sharenting has led to the victimisation of minors, there are systemic vulnerabilities in current sharenting practices that can cause the perpetration of harms (Lavorgna et al., 2022).

Sharenting has been linked to contemporary 'societal showcaseisation' (Codeluppi, 2007), the desire for visibility and the creation for money-making opportunities by parents working as social media influences (as in the case of so-called 'mumpreneurs', see Archer 2019, but also by 'Instadads' performing 'sharenting labour' as studied by Campana et al., 2020). In-depth analyses of sharenters, however, reveal a more complex reality, with many parents facing ethical dilemmas and hesitations in sharing information about their children online, as social media platforms offer a range of affordances that were once unimaginable (Chalklen and Anderson, 2017; Cino and Demozzi, 2017; Brownlie, 2018; Archer, 2019). Overall, they try to enhance their self-representational agency and ability to build community and social capital, whilst remaining mindful of surveillance and potential risks related to children's exposure in digital spaces (Chalklen and Anderson, 2017; Cino and Fomenti, 2021). At the same time, they must contend with data technologies that reinforce the cultural value of archival time and encourage users to track and document their lives (Barassi, 2020). Additionally, virtual spaces matter for parenting as places to gather experiential information, and social and practical support (Johnson, 2015; Kumar and Schoenebeck, 2015; Ranzini et al., 2020). Of course, the level of disclosure can vary, both in breadth (i.e., the amount of disclosed information, which includes the frequency and duration of disclosed contents), and in depth (which reflects the level of intimacy) (Wheeles and Grotz, 1976).

A core problem is that much of the focus of studies exploring digital parenting has been on offering guidance to adults on how to best manage screens and media for their children (Uhls, 2015). But parents (and, more generally, guardians, including teachers), whilst being expected to educate and protect children online, are unprepared for their role as 'digital custodians' (Buchanan et al., 2019: 175). They should be the first line of defence, but at times are not (Brosch, 2018). In most countries, there are currently no policies securing children's right to online privacy, and the decision on whether and how to disclose information online is left in the parent's hands. This makes them the principal gatekeepers of the personal information of their children online (Steinberg, 2017; Brosch, 2018).

It can be argued that sharenters' agency is somehow constrained by the 'systemic coercion of digital participation' at the basis of the surveillance capitalism (Barassi, 2019: 415) so embedded in our 'onlife' (Floridi, 2015; see also Barassi, 2020; Chayko, 2020). This issue can be considered as part of the 'next generation privacy problems' (Zittrain, 2008: 205), as individuals are enabled to deeply compromise privacy through the generative technologies currently in use (generative meaning their capacity and purpose is beyond what was first imagined when they were created, see Zittrain, 2008). We also need to take into consideration that children nowadays are 'datafied' in numerous ways, as they digitally participate to society without their consent or control (Lupton and Williamson, 2017; Barassi, 2019).

As already noted, most research on sharenting focus on the sharenters and their agency. The structural, regulatory characteristics of the social media platforms they inhabit, which are at the basis of their potentially harmful sharing activities, have been so far ignored. Nonetheless, these platforms' characteristics are pivotal, as they are an integral part of the digital ecosystem of interest. In what follows, we present our study which explored whether and how five mainstream social media platforms respond to potentially harmful sharenting practices. In doing so, we unravel several criminogenic and harm-enabling features of the self-regulatory strategies framing the platforms.

Methodology

Criminologists have long argued that regulations and their implementation can sometimes be counterproductively and unintentionally criminogenic (Sutherland and Cressey, 1978). In more recent years, crime proofing of legislation was developed as an approach to assess existing or future opportunities for crime due to legislation and indicate potential interventions aimed at proofing it against crime, and proved to be a valid form of risk assessment and management (Albrecht and Kilching, 2002; Russell and Clarke, 2006; Transcrime, 2006; Morganti et al., 2020). Crime proofing of legislation stems from crime opportunity approaches, sharing the core idea that opportunity is a root cause of crime (Clarke, 2012) and aiming to reduce crimes by looking for crime patterns in specific settings (Felson and Clarke, 1998) – including the regulatory ones. Crime proofing of legislation was initially developed to assess the risk of unintended consequences produced by legislative measures. It enabled policy makers to evaluate legislation in the law-making process and to suggest changes (e.g., textual changes) to reduce the crime risk (see, for instance, the MARC model in Transcrime, 2006; Savona, 2017). The approach has been adapted over the years, for instance to assist the systematic analysis of norms regulating a certain market to assess whether they might have an unintended criminogenic role, or whether there are major loopholes in their implementation that could be criminally exploited (consider, for instance, Lavorgna et al., 2018).

Our approach furthers this latter trend by using an adapted form of crime proofing of legislation as a framework to analyse aspects of the relevant text and identify in the self-regulations of social media platforms criminogenic or otherwise harm-enabling opportunities for harmful sharenting. In our revised model, we adjusted the approach used by Lavorgna and colleagues (2018, which was in itself an adaptation of Transcrime, 2006). This was necessary because we focused not only on fully-fledged crimes but also on harmful behaviours, and on existing self-regulations rather than on formal legislation under development. The *intended consequences* of the self-regulatory practice, as discussed above, are to govern and manage users' activities and relationships (among themselves and with the platform), fostering their participation whilst preventing and mitigating crimes and harms, with a focus on the legal risks they might produce. The *unintended consequences*, for the scope of our work, are the facilitation of harmful sharenting practices, or the failure in addressing them. The mismatch between the intended and the unintended consequences was assessed by looking at a series of indicators.

First, we distinguished a *regulatory risk* element (rather than a 'legislative crime threat', as in previous studies), based on the following indicators: accessibility of information (which includes accessibility in practice, and clarity of content); consistency (temporal; internal; across platforms); regulatory gaps (to what extent harmful sharenting is addressed); and implementation mechanisms (to address harmful sharenting). Second, to assess *sharenting vulnerabilities* (rather than market vulnerabilities, as in previous studies), the following indicators were considered: attractiveness (which, in our case, depends on desirability of sharing potentially sensitive information in a specific social media context), shareability and availability (respectively, how easy is to share potentially sensitive information, and the capacity to access shared material by a third party), and lack of guardianship (moderation practices at different levels).

It is worth noting that, by adapting a method derived from crime opportunity approaches, we do not want to deny, or minimise, the fact that social media use reflects, and depends on, a range of cultural, psychological, and even emotional habits, patterns and needs (e.g., Waters and Ackerman 2011; Hu et al., 2018, Stsiampkouskaya et al., 2021; Bayer et al. 2022); also, we acknowledge that most users may not read or consciously agree to platform policies. Indeed, a major limitation of this approach is that it does not fully capture how sociotechnical affordances enable and constrain behaviours (Gibson, 1977; Hutchby, 2001; Bloomfield et al., 2010). Nonetheless, as discussed above, self-regulations – by defining an important part of the relationship between social media platforms and their users, hence fabricating and delimiting our digital ecosystem of interest – remain pivotal to assessing whether and to what extent social media companies are addressing the criminogenic or otherwise harm-enabling potential of their platforms.

For our analysis, we focused on what we expected to be five social media platforms with a major role in sharenting practices. They were selected considering their demographic distribution among the population and following a discussion with our non-academic project partners[1]. The platforms are: Facebook; TikTok; Instagram; You Tube; and Twitter.

Textual data were retrieved (from the UK) from each of the platform's relevant self-regulatory documents (mostly Terms and Conditions, and Community Standards, in their English version). Some additional documents targeting, for instance, specifically parents were also identified and considered for the analysis. The data we used are publicly available online. After an initial screening, all sections of the documents that could (broadly) be relevant to sharenting, digital harms and crimes, moderation practices, minors, and their parents/guardians were selected for analysis (updated on November 30, 2021). These amounted to a total of 330 pages (108,502 words), available at [to be added after peer review]. We then performed a content analysis on those documents with the support of NVivo12, based on the following main codes, which are at the basis of the indicators mentioned above and discussed in the following section. The codes were created based on insights from emerging data and the extant literature, and they are: platform; relevant platform's metadata; platform's age limitation; provisions specific to minors (including definition); content monetisation; specific provision on someone else posting material on a subject; specific (harm/crime) risk that might occur if someone else is posting material on a subject/on behalf of a subject; crime/harm addressed; provision specific to parents/guardians/educators; sharenting if specifically addressed; content moderation practices; other. For the sake of brevity, a summary of selected descriptive results for exemplary purposes is available in Appendix A.

Identifying and discussing criminogenic features

Regulatory risk

Accessibility of information. In terms of accessibility of information, the regulations analysed are problematic. Even if they can be easily accessible online via the platforms or standard search engines by any interested party and the language used is generally clear, the relevant information is dispersed across a number of documents, sections and pages, making it hard to reach in practice. The scale of information to be navigated is extensive, making it unlikely that the standard user will be aware of relevant provisions. Parents/guardians and minors are addressed directly only in a few cases.

Consistency. Temporally, the documents analysed are relatively stable (considering the speed of innovations in the sector). Versions of the documents were last updated in the period ranging from May 2018 (Twitter User Agreement) to October 2021 (YouTube Channel Monetisation Policies and TikTok Privacy Policy). Updates, however, are not easily traceable by users, who realistically are not aware of changes in policies.

The internal of consistency of regulations (within each platform) is generally good, even if there are some provisions that could create confusion, especially when it comes to the age limitations set to use a specific platform or some of its services. These provisions are of interest in the case of sharenting practices as their presence suggests that a platform might have content inappropriate for those below a certain age but who, paradoxically, might find themselves exposed in that very same platform. On this matter, an example of internal inconsistency is provided by YouTube, where the Terms of Service states that minors under 14 can access contents only with parental guidance and with specific restrictions, but a separate section targeting specifically parents sets the access age at 13.

Consistency across platforms, not surprisingly, is far more problematic. Staying on the example of age limitations we can note differences regarding, for instance, the minimum age for subscription (generally 13 years, but 16 years for services such as TikTok Live Stream Program or Twitter Periscope). There are also disparities in policies on access to specific content, content creation, or the age under which parental control is necessary. These differences across platforms can create confusion in a context where most users operate across platforms in their digital modus operandi (Tandoc et al., 2019).

Regulatory gaps

Harmful sharenting falls within a regulatory gap in the platforms observed, with only some of its manifestations (namely, those linked to potential sexual abuse and some serious forms of antagonistic online behaviours) being taken into proper consideration in the platforms' ecosystem.

For instance, we found a number of provisions addressing the possibility of someone posting material on another subject (e.g., a child), and the provisions at times also identified specific (crime/harm) risks, through detailed lists (as in the case of Facebook and TikTok) or through more general provisions. All platforms showed attention to the issue of posting material on another subject, e.g., Instagram:

> 'You can't post someone else's private or confidential information without permission or do anything that violates someone else's rights, including intellectual property rights (e.g., copyright infringement, trademark infringement, counterfeit or pirated goods). [...] Post only your own photos and videos, and always follow the law. [...] Share only photos and videos that you've taken or have the right'.

But only some platforms had specific provisions regarding minors. None addressed sharenting in general as problematic, with platforms rather focusing on a limited number of sharenting-enabled crimes and harms.

TikTok, for example, focuses on the limitations concerning the use of material which is confidential or property of someone else, offering a detailed list of contents which can't be shared/posted (e.g., *'any material, i.e. deliberately designed to provoke or antagonise people, especially trolling and bullying, or is intended to harass, harm, hurt, scare, distress, embarrass or upset people*'). Instagram offers (less detailed) provisions focusing on the need to avoid abusive behaviours (e.g., sharing nude or sexual photos/videos is considered violation of Instagram's Community Guidelines). But minors are never specifically mentioned in this context. Facebook, on the other hand, provides several specific provisions about avoiding posting material that could led to bullying and harassment (e.g., by forbidding '*content manipulated to highlight, circle or otherwise negatively draw attention to specific physical characteristics […]*'), or the sexual exploitation of children. Most of these previsions are then specified in a detailed way (e.g., the topic of non-sexual child abuse is addressed by considering imagery posted by news agencies depicting children in sensitive contexts). However, specific alerts concerning the 'mere' posting of textual or visual information on minors by parents or guardians are never addressed as a specific matter of concern, with the exception of the potential for sexual exploitation of children through the sharing of images. Statements illustrating this include:

> 'We do not allow content that sexually exploits or endangers children. […] We know that sometimes, people share nude images of their own children with good intentions; however, we generally remove these images because of the potential for abuse by others and to help avoid the possibility of other people re-using or misappropriating the images'.

There are some provisions specific to parents, guardians and educators, but also in these cases sharenting is not addressed in its entirety. For instance, TikTok offers detailed provisions to parents/guardians/educators through its Terms of Service and a Guardians' Guide. However, this information seems to address how teenagers use TikTok, and not issues concerning sharenting. In other words, the orientation that TikTok documents offer is useful to guide guardians in supervising or monitoring minors actions on the platform. But it does not mention directly the sharing practices of these guardians which instead are ruled by the general terms of use for every adult users. Similarly, YouTube's guidance to guardians is limited to the control of the kind of contents that children can watch, the time they spend on the platform, and similar issues. Thus, the platform highlights the responsibility of parents and guardians in monitoring children's activities on the platform. Compared to the other platforms, Instagram appears to address sharenting more explicitly, but again the focus is mostly limited to the potential for sexual exploitation:

> 'People like to share photos or videos of their children. For safety reasons, there are times when we may remove images that show nude or partially nude children. Even when this content is shared with good intentions, it could be used by others in unanticipated ways. You can learn more on our Tips for parents page'.

Implementation mechanisms

In order to intervene in existing illicit conducts and to prevent future ones, all social media platforms have a number of implementation and enforcement mechanisms which include, among other things: reporting tools; blocking and filtering; geofencing (i.e., banning a certain geographical location to access a social media platform);

various forms of manual or automated moderation; educational material (Milosevic, 2017). With reference to the provisions explored to far, overall, in all five platforms, implementation mechanisms are primarily based on some form of collective responsibility at community level involving the monitoring of material posted, and reporting it if and when appropriate (e.g., on YouTube: '*If you see a video that you feel is inappropriate, flag the video*'). As such, the reporting is delegated to the community (similar to what was noted by Milosevic in regards to cyberbullying) – as if it is 'primarily the responsibility of the community to regulate itself rather than have the company regulate the community' (Milosevic, 2017: 114). Additionally, posts are then removed if their content is among those listed as forbidden. As such, harmful sharenting is not addressed through specific enforcement mechanisms, with the exception of its more extreme manifestations as discussed above. Similarly, compliance with other requirements issued by legislators (such as age requirements verification) is not realised proactively. An example is the apparent neglect of age limits which are only enforced when reported by other users as already stressed by Schneble and colleagues (2021).

It is worth noting that, despite their presence on the platforms considered, minors are relatively disregarded as agentic users in the documents analysed, and rather addressed as an audience to be protected (e.g., YouTube: '*You're required to tell us that your videos are made for kids if you make kids content. As a YouTube creator, you are required to set future and existing videos as made for kids or not*'). Specific provisions addressing minors generally focus on age limitations (as discussed above). In the case of Facebook there is a provision specifying that minors between 13 and 18 can report content if they notice a misuse of their images or videos. As such, according to Facebook, the following content also may be removed: '*A reported photo or video of people where the person depicted in the image is: A minor under the age of 13, and the content was reported by the minor or a parent or legal guardian. A minor between the ages of 13 and 18 years old, and the content was reported by the minor*'. Instagram seems to provide more attention to the topic, addressing minors directly, for instance, encouraging them to report bullying and harassment through its Help Center, and to reflect on the opportunity to share nude or sexual photos or videos.

Sharenting vulnerabilities

Attractiveness. The sharing of potentially sensitive information of minors can originate from motivations ranging from parental pride and social isolation, to the search for social influence and profit (Lavorgna et al., 2022). Whilst some of these factors are inherent in the nature of social media platforms (Haillikainen, 2015), some can directly increase the rewards for harmful sharenting practice. Consider, in particular, the monetisation of certain content. TikTok and YouTube, for instance, provide specific self-regulatory provisions about this issue, with a few references to minors and age limitations, but again, with minors considered as potential audience and not as potential 'content'. For example, in defining contents which are not suitable for monetisation (e.g., '*content that incites hatred against, promotes discrimination, disparages, or humiliates an individual or group of people*'), there is no specific provision for protecting the exposed identities of minors.

Shareability and availability

In all the platforms considered, it is extremely easy for anyone enrolled to the platform to share potentially sensitive information of minors; there are no awareness raising tools or other mechanisms intervening, unless the content is manifestly forbidden as discussed above. Similarly, it is easy for a third party to access (and re-share) the material posted, with limited risk of detection. It is possible to set the privacy settings of both individual accounts and groups/pages to limit these possibilities, but these are not the default options.

Lack of guardianship

Relevant content moderation practices can occur at both platform and community/individual levels. Platform moderation generally takes place in automatised ways, with human moderators being kept 'in the loop' and intervening when needed. The 'logic of opacity' (Roberts, 2018) previously discussed is found also in this context, as the details of the training (of software; of human beings) informing moderation practices are not publicly available, and the information available on the standards applied to moderation is obfuscated. This opacity by design affects standard moderation and reporting practices, as well atypical procedures envisaged by some platforms, such as expert moderation or social reporting (e.g. on Facebook). It is expected that the latter in particular will be used alongside the regular reporting to try to resolve issues not listed by corporate policy. The procedures are accessible through a series of prompts whose wording is based on the user's characteristics. Such procedures might have the potential to counter digital harms, but they have to be regularly and independently evaluated (as discussed also in Milosevic, 2017: 122ff).

For the more serious manifestations of sharenting addressed by the platforms, there are actions to protect minors at the platform level. Consider, for instance, the following snippet from Facebook Community Standards:

> 'We have built a combination of automated and manual systems to block and remove accounts that are used to persistently or egregiously abuse our Community Standards. […] We comply with: Requests for removal of an underage account. Government requests for removal of child abuse imagery depicting, for example, beating by an adult or strangling or suffocating by an adult. Legal guardian requests for removal of attacks on unintentionally famous minors. […] We may remove content created for the purpose of identifying a private minor if there may be a risk to the minor's safety when requested by a user, government, law enforcement or external child safety experts'.

But, again, there are no instruments to intervene in less extreme yet potentially harmful cases. As such, should an individual signal the presence of a potentially inappropriate post, which for instance, discusses some sensitive information concerning a minor, there is no clear remedy.

Here, also the discretionary power of the platforms is evident, not only because of the use of modal verbs such as 'may', but also because of the platforms' ability to unilaterally decide whether and how to intervene. On this aspect, consider also the following excerpt from TikTok Terms of Service:

> '[…] We have the right to remove, disallow, block or delete any posting you make on our Platform if, in our opinion, your post does not comply with the content standards set out at *Section 7* (Your Use of Our Services) above. In addition, we have the right – but not the obligation – in our sole discretion to remove, disallow, block or delete any User Content (i) that we consider to violate these Terms, or (ii) in response to complaints from other users or third parties, with or without notice and without any liability to you. […]'.

Instagram and YouTube offer a more balanced approach between the several moderation levels, requiring that the users become active in reporting harmful material, and even reminding users of their agency in controlling digital abuse. Consider, for instance, this snippet form YouTube Community Guidelines (Reporting and Enforcement):

> 'We rely on YouTube community members to report, or flag content that they find inappropriate. Reporting content is anonymous, so other users can't tell who made the report. […] When something is reported, it's not automatically taken down. Reported content is reviewed along the following guidelines: Content that violates our Community Guidelines is removed from YouTube. Content that may not be appropriate for younger audiences may be age-restricted. […] If you see a video that you feel is inappropriate or which may violate our Community Guidelines, flag the video. […] YouTube policy specialists review flagged videos 24 h a day, 7 days a week. […] If you feel that your child's privacy has been violated (e.g., use of image or personal information without consent), please visit our Privacy Guidelines, where you can learn more about our privacy policy and how to file a privacy complaint'.

Twitter, on the other hand, emphasises the role of community moderation, by acknowledging the limits of its ability to automatically intervene from a platform's level ('*We may not monitor or control the Content posted via the Services and, we cannot take responsibility for such Content*'). This suggests that, even if Twitter can take actions against violations, the responsibility to respect the rules is assigned to the users.

Further discussion and conclusion

We have discussed the regulation and moderation instituted by major social media companies which, through their self-regulations, set the boundaries of what contents and behaviours are allowed on their platforms. We argue that unpacking the self-regulatory strategies is imperative to deepen our understanding of their dynamics, rhetoric, and effectiveness (in line with Wyatt, 2008). This is necessary since they define the digital ecosystem formed by the relationship between social media platforms and their users. As such, they frame users' perceptions and practices in a climate of increasing privatization of the digital public sphere (Milosevic, 2017: 47).

In the context of 'deep mediatization' (Hepp, 2020) permeating our social spaces, more and more aspects of our lives – including how we frame and manage risks – are seamlessly adapted to digital media logic and infrastructures. But even if users maintain their agentic capacities (Lupton, 2020; Lupton and Sutherton, 2021), their ability to act is both enabled and constrained by (technological and regulatory) social media affordances (e.g., Gibson, 1977; Hutchby, 2001; Bloomfield et al., 2010). By shaping conditions of possibility, and hence overcoming the limitations of technological determinism (Fussey and Roth, 2020), social media affordances leave space for human (individual or collective) choice or intervention. In the process, they hold us responsible for the technologies we make and use (Wyatt, 2008). Individual and collective users' responsibility for harmful sharenting practices cannot, and should not, be minimised. Nonetheless, as evidenced in our findings, these practices can be enabled and even facilitated by a defective self-regulatory framework: by using an adapted crime proofing of legislation approach, we highlighted severe regulatory gaps, and how some attempts to regulate user behaviour might instead give rise to increased risk. Users such as parents/guardians sharing minors' information, and their audiences and secondary distributors, have been defined as 'slack and irresponsible' (Lim et al., 2020: 96) for creating criminogenic opportunities online. We do acknowledge that our analysis of policy documents and guidelines may not fully capture the myriad of ways in which the lack of adequate policy we observed provides opportunities for illegal or otherwise harmful sharenting practices or how users may experience the inadequate policies instituted by platforms. Our findings demonstrate that we should not ignore the platforms' responsibilities – even if the deniability of their legal responsibility has been at the core of the rise of contemporary digital giants (Gillespie, 2010; Andersson Schwartz, 2017; Klonick, 2017; Gorwa, 2019).

Of course, regulating the power of platforms is not an easy task (Lynskey, 2017; Bucher, 2018), and we firmly agree with the importance of self-regulation and its implementation through moderation systems. Yet, social media design should more carefully consider the impact of content moderation in practice on potentially harmful behaviour, as the mere presence of platform regulations does not guarantee efficient and effective practices (Chancellor et al., 2016).

From a criminological standpoint, having identified specific criminogenic features, Situational Crime Prevention techniques and mechanisms (see Clarke, 1992, 2009; Freilich and Newman, 2014) can provide useful guidance on how to mitigate the problems identified. The techniques offer a framework for mapping and devising potential interventions. Such interventions could be useful for addressing harmful sharenting practices. The interventions could also help address other comparable digital activities that are potentially harmful in the sense that they can harm individuals in ways that are currently escaping existing self-regulatory approaches. Discussing practical solutions would exceed the scope of this study. But our contribution has shown how criminological imagination can effectively contribute to multidisciplinary debates not only on sharenting practices but, more broadly, on digital ecosystems and their regulation, paving the way for a reduction of criminogenic and harming opportunities online.

Acknowledgments

We wish to thank the anonymous reviewers for their very constructive feedback.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by the UK ESRC, project 'ProTechThem: Building Awareness for Safer and Technology-Savvy Sharenting', ES/V011278/1.

ORCID iD

Anita Lavorgna https://orcid.org/0000-0001-8484-1613

Note

1. See https://www.protechthem.org/people/ for details.

References

Abbott KW and Snidal D (2009) The governance triangle: Regulatory standards institutions and the shadow of the state. In: Mattli W and Woods N (eds) *The Politics of Global Regulation*. Princeton, NJ: Princeton University Press, pp.44–88.

Albrecht HJ and Kilching M (2002) Crime risk assessment, legislation, and the prevention of serious crime. Comparative perspectives. *European Journal of Crime, Criminal Law and Criminal Justice* 10(1): 25–38.

Ammari T, Kumar P, Lampe C, et al. (2015) Managing children's online identities: How parents decide what to disclose about their children online. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Seoul: ACM, 1895–1904.

Andersson Schwarz J (2017) Platform logic: An interdisciplinary approach to the platform-based economy. *Policy & Internet* 9(4): 374–394.

Archer C (2019) How influencer 'mumpreneur' bloggers and 'everyday' mums frame presenting their children online. *Media International Australia* 170(1): 47–56.

Barassi V (2019) Datafied citizens in the age of coerced digital participation. *Sociological Research Online* 24(3): 414–429.

Barassi V (2020) Datafied times: Surveillance capitalism, data technologies and the social construction of time in family life. *New Media & Society* 22(9): 1545–1560.

Barnes R and Potter A (2021) Sharenting and parents' digital literacy: An agenda for future research. *Communication Research and Practice* 7(1): 6–20.

Bayer JB, Anderson IA and Tokunaga RS (2022) Building and breaking social media habits. *Current Opinion in Psychology* 45: 101303.

Bezáková Z, Madleňák A and Švec M (2021) Security risks of sharing content based on minors by their family members on social media in times of technology interference. *Media Literacy and Academic Research* 4: 53–69.

Bloomfield BP, Latham V and Vurdubakis T (2010) Bodies, technologies and action possibilities: When is an affordance? *Sociology* 44(3): 415–433.

Brosch A (2018) Sharenting: Why do parents violate their children's privacy? *The New Educational Review* 54(4): 75–85.

Brownlie J (2018) Looking out for each other online: Digital outreach, emotional surveillance and safe(r) spaces. *Emotion, Space and Society* 27: 60–67.

Buchanan Southgate E and Smith SP (2019) 'The whole world's watching really': Parental and educator perspectives on managing children's digital lives. *Global Studies of Childhood* 9(2): 167–180.

Bucher T (2018) *If... Then: Algorithmic Power and Politics*. Oxford: Oxford University Press.

Campana M, Van den Bossche A and Miller B (2020) #Dadtribe: Performing sharenting labour to commercialise involved fatherhood. *Journal of Macromarketing* 40(4): 475–491.

Chalklen C and Anderson H (2017) Mothering on Facebook: Exploring the privacy/openness paradox. *Social Media & Society* 3: 205630511770718.

Chancellor S, Pater J, Clear T, et al. (2016) #thyghgapp: Instagram Content Moderation and Lexical Variation in Pro-Eating Disorder Communities. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, pp.1201–1213.

Chayko M (2020) *Superconnected: The internet, digital media, and techno-social life*. Thousand Oaks, CA: Sage.

Choi GY and Lewallen J (2018) 'Say Instagram, Kids!': Examining sharenting and children's digital representations on Instagram. *Howard Journal of Communications* 29(2): 144–164.

Cino D and Demozzi S (2017) Figli 'in vetrina'. Il fenomeno dello sharenting in un'indagine esplorativa. *Rivista Italiana di Educazione Familiare* 2: 153–184.

Cino D and Formenti L (2021) To share or not to share? That is the (social media) dilemma. Expectant mothers questioning and making sense of performing pregnancy on social media. *Convergence* 27(2): 491–507.

Clarke RV (1992) *Situational crime prevention: Successful case studies*. New York: Harrow and Heston.

Clarke RV (2009) Situational Crime Prevention. In: Wortley R and Mazerolle L (eds) *Environmental criminology and crime analysis*. Devon: Willan, pp.286–303.

Clarke RV (2012) Opportunity makes the thief. Really? And so what? *Crime Science* 1(2): 1–9. doi:10.1186/2193-7680-1-3.

Codeluppi V (2007) *La vetrinizzazione sociale. Il processo di spettacolarizzazione degli individui e della società*. Turin: Bollati e Boringheti.

Cusumano MA, Gawer A and Yoffie DB (2021) Can self-regulation save digital platforms? *Industrial and Corporate Change* 30(5): 1259–1285.

DCMS (2019) Disinformation and "fake news" [Report No. 8]. Digital, Culture, Media and Sport Committee. London: House of Commons.

Di Bari C (2017) L'infanzia rappresentata dai genitori nei social network: Riflessioni pedagogiche sullo sharenting. *Studi sulla Formazionw*: 257–271.

Douek E (2019) Facebook's 'Oversight Board': Move Fast with Stable Infrastructure and Humility. *N.C. J. L. & Tech* 21(1): 1–77.

Escobar A (1994) Welcome to cyberia: Notes on the anthropology of cyberculture. *Current Anthropology* 35: 211–231.

Felson M and Clarke RV (1998) Opportunity makes the thief: practical theory for crime prevention. Police Research Series 98, London: Home Office.

Floridi L (2015) *The Onlife Manifesto: Being Human in a Hyperconnected Era*. Cham: Springer.

Freilich JD and Newman GR (2014) Providing opportunities: A sixth column for the techniques of situational crime prevention. In: Caneppele S and Calderoni F (eds) *Organised crime, corruption and crime prevention*. London: Springer, pp.33–42.

Fussey P and Roth S (2020) Digitizing sociology: Continuity and change in the internet era. *Sociology* 54(4): 659–674.

Gibson J (1977) The theory of affordances. In: Shaw R and Bransford J (eds) *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Hillsdale, NJ: Erlbaum, pp.67–82.

Gillespie T (2010) The politics of 'platforms'. *New Media & Society* 12(3): 347–364.

Gillespie T (2017) Governance of and by platforms. In: Burgess J, Poell T and Marwick A (eds) *The SAGE Handbook of Social Media*. New York: Sage, pp.254–278.

Gillespie T (2018) *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*. New Haven, CT: Yale University Press.

Gorwa R (2019) The platform governance triangle: Conceptualising the informal regulation of online content. *Internet Policy Review* 8(2). https://doi.org/10.14763/2019.2.1407.

Gorwa R, Binns R and Katzenbach C (2020) Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7: 205395171989794.

Grimmelmann J (2015) The virtues of moderation. *Yale Journal of Law & Technology* 17: 42.

Hallikainen P (2015) Why People Use Social Media Platforms: Exploring the Motivations and Consequences of Use. In: Mola L, Pennarola F and Za S (eds) *From Information to Smart Society. Lecture Notes in Information Systems and Organisation, vol 5*. Cham: Springer, pp.9–17.

Hancock H (2021) The impact of the image on personal life: Is current law out of focus? *Journal of Media Law* 13(1): 54–80.

Hepp A (2020) *Deep Mediatization*. New York: Routledge.

Hu T, Stafford TF, Kettinger WJ, et al. (2018) Formation and effect of social Media usage habit. *Journal of Computer Information Systems* 58(4): 334–343.

Hutchby I (2001) Technologies, texts and affordances. *Sociology* 35(2): 441–456.

Johnson BG (2017) Speech, harm, and the duties of digital intermediaries: Conceptualizing platform ethics. *Journal of Media Ethics* 32: 16–27.

Johnson SA (2015) 'Intimate mothering publics': Comparing face-to-face support groups and internet use for women seeking information and advice in the transition to first-time motherhood. *Culture, Health & Sexuality* 17(2): 237–251.

Keller D and Leerssen P (2019) Facts and Where to Find Them: Empirical Research on Internet Platforms and Content Moderation. In: Persily N and Tucker J (eds) *Social Media and Democracy: The State of the Field and Prospects for Reform*. Cambridge: Cambridge University Press, pp.220–250.

Klonick K (2017) The new governors: The people, rules, and processes governing online speech. *Harvard Law Review* 131(6): 1598–1670.

Kumar P and Schoenebeck S (2015) The modern day baby book: Enacting good mothering and stewarding privacy on Facebook. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, ACM, 1302-1312.

Lavorgna A (2021) Looking and crime and deviancy in cyberspace through the social harm lens. In: Leighton PS, Wyatt T and Davies P (eds) *Handbook of Social Harm*. Palgrave, pp.401–420.

Lavorgna A, Rutherford C, Vaglica V, et al. (2018) CITES, wild plant, and opportunities for crime. *European Journal of Criminal Policy and Research* 24(3): 269–288.

Lavorgna A, Ugwudike P and Tartari M (2022) Online sharenting: identifying existing vulnerabilities and demystifying media reported risks (under review).

Lazard L, Capdevila R, Dann C, et al. (2019) Sharenting: Pride, affect and the day-to-day politics of digital mothering. *Social and Personality Psychology Compass* 13(49): e12443.

Leerssen P (2015) Cut Out by the Middleman: The Free Speech Implications of Social Network Blocking and Banning In The EU. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 6(2).

Li S and Williams J (2018) Despite What Zuckerberg's Testimony May Imply, AI Cannot Save Us. Electronic Frontier Foundation Deeplinks Blog. Available at: https://www.eff.org/deeplinks/2018/04/despite-what-zuckerbergs-testimony-may-imply-ai-cannot-save-us.

Lim Y, Lim CM, Gan KH, et al. (2020) Text Sentiment Analysis on Twitter to Identify Positive or Negative Context in Addressing Inept Regulations on Social Media Platforms. 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE), 96-101.

Lupton D (2020) Thinking with care about personal data profiling: A more-than-human approach. *International Journal of Communication* 14: 3165–3183.

Lupton D and Sutherton C (2021) The thing-power of the Facebook assemblage: Why do users stay on the platform? *Journal of Sociology* 57: 969–985.

Lupton D and Williamson B (2017) The datafied child: The dataveillance of children and implications for their rights. *New Media and Society* 19(5): 780–794.

Lynskey O (2017) Regulating 'Platform Power' [Working Paper No.1]. London: London School of Economics and Political Science, Law Department.

Milosavljević M and Micova SB (2016) Banning, blocking and boosting: Twitter's solo-regulation of expression. *Medijske Studije* 7(13): 43–58.

Milosevic T (2017) *Protecting children online? Cyberbullying policies of social media companies*. Cambridge, MA: MIT.

Morganti M, Favarin S and Andreatta D (2020) Illicit waste trafficking and loopholes in the European and Italian legislation. *European Journal of Criminal Policy and Research* 26: 105–133.

Powell A, Stratton G and Cameron R (2018) *Digital criminology. Crime and justice in digital society*. London: Routledge.

Prochazka O (2019) Making sense of Facebook's content moderation: A posthumanist perspective on communicative competence and internet memes. *Signs and Society* 7(3): 362–397.

Ranzini G, Newlands G and Lutz C (2020) Sharenting, peer influence, and privacy concerns: A study on the instagram-sharing behaviors of parents in the United Kingdom. *Social Media & Society* 6: 205630512097837.

Roberts ST (2018) Digital detritus: 'Error' and the logic of opacity in social media content moderation. *First Monday* 23(3-5).

Russell M and Clarke RV (2006) Legislation and unintended consequences for crime. *European Journal of Criminal Policy and Research* 12: 189–211.

Santos Rutschman A (2021) *Social Media Self-Regulation and the Rise of Vaccine Misinformation*. Available at: https://scholarship.law.slu.edu/cgi/ viewcontent.cgi?article = 1553&context = faculty.

Savona EU (2017) Proofing Legislation Against Crime as Situational Prevention Measure. In: Leclerc B and Savona EU (eds) *Crime Prevention in the 21st Century*. Cham: Springer, pp.247–274.

Schneble CO, Favaretto M, Elger BS, et al. (2021) Social media terms and conditions and informed consent from children: Ethical analysis. *JMIR Pediatrics and Parenting* 4: e22281.

Seering J (2020) Reconsidering self-moderation: The role of research in supporting community-based models for online content moderation. *Proceedings of the ACM on Human-Computer Interaction* 4: 1–28.

Steinberg SB (2017) Sharenting: Children's privacy in the age of social Media. *Emory Law Journal* 839(66): 839–884.

Stsiampkouskaya K, Joinson A, Piwek L, et al. (2021) Emotional responses to likes and comments regulate posting frequency and content change behaviour on social media: An experimental study and mediation model. *Computers in Human Behavior* 124: 106940.

Sutherland EH and Cressey D (1978) *Criminology*. Philadelphia, PA: JB Lippincott.

Tandoc EC, Lou C and Lee Hui Min V (2019) Platform-swinging in a poly-social-media context: How and why users navigate multiple social media platforms. *Journal of Computer-Mediated Communication* 24(1): 21–35.

Transcrime (2006) *A study on crime proofing – Evaluation of crime risk implications of the European Commission's proposal covering a range of policy areas*. Available at: http://www.transcrime.it/wpcontent/uploads/2013/11/Final_Manual-A_study_on_Crime_Proofing.pdf.

Tyler T, Katsaros M, Venkatesh S, et al. (2021) Social media governance: Can social media companies motivate voluntary rule following behaviour among their users? *Journal Of Experimental Criminology* 17(1): 109–127.

Uhls Y (2015) *Media Mums and Digital Dads: A fact-not-fear approach to parenting in the digital age*. London: Routledge.

Wachs S, Mazzone A, Milosevic T, et al. (2021) Online correlates of cyberhate involvement among young people from ten European countries: An application of the routine activity and problem behaviour theory. *Computers in Human Behavior* 123: 106872.

Waters S and Ackerman J (2011) Exploring privacy management on Facebook: Motivations and perceived consequences of voluntary disclosure. *Journal of Computer-Mediated Communication* 17(1): 101–115.

West SM (2018) Censored, suspended, shadow banned: User interpretations of content moderation on social media platforms. *New Media Soc* 20: 4366–4383.

Wheeless LR and Grotz J (1976) Conceptualization and measurement of reported self-disclosure. *Human Communication Research* 2(4): 338–346.

Williams-Ceci S, Grose GE, Pinch AC, et al. (2021) Combating sharenting: Interventions to alter parents' attitudes toward posting about their children online. *Computers in Human Behavior* 125: 106939.

Wise J (2019) Social media: end self-regulation, say MPs in report on children's health. *British Medical Journal* 364: 1486–1489.

Wyatt S (2008) Technological determinism is dead; long live technological determinism. In: Hackett E, Amsterdamska O, Lynch M and Wajcman J (eds) *Handbook of Science and Technology Studies*. pp.165–180.

Zankova B and Dimitrov V (2020) Social Media Regulation: Models and Proposals. *Journalism and mass communication* 10: 75–88.

Zittrain J (2008) *The future of the Internet and how to stop it*. New Haven: Yale University Press.

Appendix

A Codebook and selected samples

Code	Platform	Document	Text or research notes (selected sample)

Platform	Facebook: Terms of Service and Community Standards;
	Instagram: Terms of Use, Community Guidelines; Guide for Parents;
	Twitter: User Agreement (incl. Terms of Service, Rules and Policies, and Privacy Policy);
	TikTok: Terms of Service, Community Guidelines, Guardians’ Guide;
	YouTube: Terms of Service, Community Guidelines		

Relevant platform’s metadata	Date of last revision (sample):
	Facebook Terms of Service: 20 December 2020;
	Facebook Community Standards: 21.11.2021
	Instagram Terms of Use: 20 December 2020;
	Instagram Community Guidelines FAQs: 19 April 2018;
	TikTok Terms of Service: July 2020;
	TikTok Privacy Policy for younger users: January 2020;
	TikTok Community Guidelines: December 2020
	Twitter Terms of Service and Privacy Policy: 25 May 2018;
	YouTube Terms of Service: 01.06.2021;
	YouTube channel monetization policies: October 2021		

Age limitation to use the platform	Facebook	Terms of Service	‘You cannot use Facebook if: you are under 13 years old’.
	Instagram	Terms of Use	Minors under 13 cannot subscribe as Instagram users.
	TikTok	Terms of Service	Minors under 13 cannot subscribe as TikTok users. Minors under 16 cannot access TikTok Live Stream Program: they need parents or legal guardian’s full permission if between 16 and 18. ‘The TikTok experience for Younger Users’ is only for users under 13 in the US. Minors under 18 cannot use TikTok Careers
		Community Guidelines,	‘If someone is sharing photos or videos that make you

(Continued)

Appendix. (Continued)

Code	Platform	Document	Text or research notes (selected sample)
Provision specific to minors (incl. definition)	Instagram TikTok	Safety tips, Tips for parents Community Guidelines	uncomfortable, you can unfollow or block them. You can also report something that you feel violates our Community Guidelines right from the app'. 'By default, accounts for people under 16 are set to private which means you can approve or deny follower requests, and only people you've approved as followers can see your content. Accounts for people who are over 167 start out as public, which means any TikTok community member can view your public videos and post comments to engage with the content you've created and shared'.
Content monetisation	YouTube TikTok	YouTube channel monetisation policies Virtual Items Policy	'Starting in November [2021], the quality principles for kids and family content will be used to make monetization decisions for content classified as "made for kids"'. Users can only buy Coins and Gifts, send Gifts to others, receive Gifts with monetary value, earn Diamonds and withdraw Diamonds if they are aged 18 (or age of majority in your jurisdiction) or older. […]
Specific provision on someone else posting material on a subject	Facebook TikTok	Community Standards Terms of Service	'Do not post: Content that threatens, depicts, praises, supports, provides instructions for, makes statements of intent, admits participation in or shares links of the sexual exploitation of children (real or non-real minors, toddlers or babies), including, but not limited to: Sexual intercourse […] Children with sexual elements […]' You must not post any User Content on or through the Services or transmit to us any User Content that you consider to be confidential or proprietary to any other person.
Specific (crime/harm) risk that might occur			'But beyond Instagram, it's important to think about how it

(Continued)

Appendix. (Continued)

Code	Platform	Document	Text or research notes (selected sample)
if someone is posting material on a subject/on behalf of a subject	Instagram Facebook	Safety Tips Community Standards	would impact you if nude or sexual photos or videos of yourself got out of your control. This can happen the instant someone shares a photo or video. Sometimes people make mistakes, play stupid jokes or act in anger and share things they shouldn't.' Do not: Post content that targets private individuals through unwanted Pages, groups and events. We remove this content when it is reported by the victim or an authorised representative of the victim.
Crime/harm addressed	TikTok Facebook	Community Guidelines Community Standards	'Sexual exploitation of minors is defined as any abuse of a position of power or trust for sexual purposes, including profiting financially, socially, sexually, or politically from the exploitation of a minor. Child Sexual Abuse Material (CSAM) is defined as any visual depiction of sexually explicit nudity or conduct, whether captured by predatory adults, peers, or self-generated by minors. TikTok will take action on any content or accounts involving sexual interactions and advances between an adult and a minor, or between minors with a significant age difference'. 'For minors: Comparisons to animals or insects that are culturally perceived as intellectually or physically inferior or to an inanimate object ("cow", "monkey", "potato"). Content manipulated to highlight, circle or otherwise negatively draw attention to specific physical characteristics (nose, ear and so on)'.
Provision specific to parents/ guardians/ educators	YouTube TikTok	Terms of Service Guardians' Guide	'If you are a parent or legal guardian of a user under the age of 18, by allowing your child to use the Service, you are

(Continued)

Appendix. (Continued)

Code	Platform	Document	Text or research notes (selected sample)
			subject to the terms of this Agreement and responsible for your child's activity on the Service. […] You can find tools and resources to help you manage your family's experience on YouTube (including how to enable a child under the age of 14 to use the Service and YouTube Kids) in our Help Center and through Google's Family Link'. 'Hosting candid family conversations about the ways in which your teen engages online, including and beyond TikTok, will help bolster their sense of digital citizenship and empower them to be mindful of their own safety on the internet'.
Sharenting if specifically addressed	Facebook	Community Standards	'We know that sometimes, people share nude images of their own children with good intentions; however, we generally remove these images because of the potential for abuse by others and to help avoid the possibility of other people re-using or misappropriating the images'.
Content moderation practices	Instagram Facebook	Policies and reporting / Tips for parents Community Standards	'[…]If your teen encounters someone on Instagram who's not following these guidelines or Terms, they can report that person's posts directly from Instagram. People can report abusive behavior or posts with our built-in reporting. This includes nude photos, abuse and excessive spam. Reporting is totally anonymous. No information about the reporter is sent to the person whose account or photo has been reported'. 'In order to maintain a safe environment and empower free expression, we also remove accounts that are harmful to the community, including those that compromise the security of other accounts and our

(Continued)

Appendix. (Continued)

Code	Platform	Document	Text or research notes (selected sample)
			services. We have built a combination of automated and manual systems to block and remove accounts that are used to persistently or egregiously abuse our Community Standards'.
Other	Instagram TikTok	Community Guidelines Law Enforcement Guidelines	'In some cases, we allow content for public awareness which would otherwise go against our Community Guidelines – if it is newsworthy and in the public interest We only do this after weighing the public interest value against the risk of harm and we look to international human rights standards to make these judgments'. 'All requests to TikTok and any supporting documents (including orders, warrants, and/or subpoenas or equivalent) must be provided in English or with a translation'.