Online safety tools — a false hope?

Is the Online Safety Bill's reliance on user-empowerment tools too complacent?

parentzone

Introduction

The Online Safety Bill will necessitate the deployment of various kinds of user-empowerment tools to fulfil its promise to keep children safe online, as well as to curate adults' feeds to reduce the likelihood of encountering harms.

These tools are often discussed as though they were a silver bullet, with no further debate required. The notion sounds logical enough: use technology to mitigate the harms technology has created. But the harms are not created by technology alone, and it is unlikely that we can simply assign safety to tools and have done with it.

Parent Zone has a number of concerns about undue reliance on tools, which we set out in this paper. Our reservations fall into three broad areas:

- How easy are the tools to use in practice?
- How effective are they?
- What are the implications for data collection and privacy?

At the time of writing, the Online Safety Bill is passing through the House of Lords, where it is subject to a number of amendments. Much about the new regulatory regime will depend on Ofcom's Codes of Practice, about which there is currently very little information. We cannot know what new tools tech companies are developing. Much is still unclear, in other words; but Parent Zone wanted to publish this paper now while there remains an opportunity for policymakers to discuss the usefulness and limitations of tools, and to guard against assumptions that the hard questions can simply be handed back to developers.

Search and social media platforms – those parts of the internet subject to regulation under the Online Safety Bill – are individually massively complex systems. The government's own estimates are that between 25,100 and 180,000 different platforms will be subject to the new provisions[1]. They are unlikely to be susceptible to one-size-fits-all solutions. Parent Zone accepts that, of course, tools have a part to play, but they need to be designed and developed with users front and centre; and their effectiveness must be understood as dependent on the circumstances, values, capabilities and media literacy of users.

[1] https://www.mishcon.com/news/the-online-safety-bill-who-is-in-potential-scope-of-the-proposed-regime#:~:text=That%20said%2C%20the%20Government%20has,in%20scope%20of%20the%20Bill.

A note on age-gating the internet

Before we come on to our reservations about the practicality of tools, it is worth noting that Parent Zone has some concerns in principle about age-gating the internet. How age assurance will work in practice will again depend to a large degree on Ofcom, in particular on how the regulator decides to reflect young people's evolving capacity. If the system is to work, it seems likely that there will have to be some granularity and parental involvement between the ages of 13 and 18 – which means more parental supervisory tools.

We have three broad concerns about age-gating:

1. **Over-blocking.** More than one platform has blocked LGBT+ content from young people in past attempts to be more child-friendly. Rather than be caught out, platforms may feel incentivised towards a 'better safe than sorry' approach, thus denying children their rights under the UNCRC, notably to information, education, and participation.
2. **Workarounds.** Research has found that children and young people are adept at finding ways to bypass restrictions. Age assurance technology may become more foolproof, but if restrictions are seen to be too draconian, it's likely that tech-savvy children will find ways to bypass them, using Virtual Personal Networks – VPNS; with fake accounts; by gaming the algorithms, and so on. Research last year by City, University of London, found that almost half (46%) of 16- and 17-year-olds were using VPNs or Tor browsers to access pornography[2].
3. **Cliff-edge at 18.** Unless there is real granularity and developmentally appropriate access, young people will be thrust out of the walled garden of the children's internet at 18 into a world of minimal regulation. (Imagine being the youngest in your cohort, or the oldest: it sounds pretty lonely.) In giving up on a duty of care to all users, policymakers have decided that the internet that young adults enter at 18 will be as toxic as ever. Young people will face the shock of an online world for which they have not been prepared – and they will have to cope with the toxicity without having had the opportunity to develop resilience, which is not by any stretch of the imagination how good parenting works. This is one reason why so many are arguing that tools blocking toxic content should be default-on.

[2] https://www.city.ac.uk/news-and-events/news/2021/05/four-in-five-uk-16-and-17-year-olds-have-seen-online-pornography-most-commonly-on-the-day-of-the-survey

How easy are parental supervisory and user-empowerment tools to use in practice?

> *'It's taken me a lot of time playing around with the apps to get to grips with the safety tools. Some are better than others.'*

A plethora of tools already exist on the platforms and services that fall in scope of the Online Safety Bill, both for parents and for adult users in general. In our discussions with parents, we often hear that they find tools confusing. They don't always know what's out there.

> *'It's difficult to keep track as a parent, but kids find them easy to use.'*

In Spring 2023, we found 33 different user-empowerment tools on Facebook alone. Some had overlapping functions or offered the same functionality: for example, it is possible to stop autoplay (where videos don't automatically play when you scroll over them) in two different places on the site. It took our twentysomething, heavy-social-media-using researcher three-and-a-half hours to find and activate all these tools. Some were easier to locate than others: she reported that finding where to deactivate and delete your account was 'unnecessarily difficult'.

> *'They're tricky to find, and they're not standardised.'*

We explored the tools on the Google search engine; on Facebook; Instagram; TikTok; Snapchat; YouTube; YouTube Kids; and WhatsApp. Between them they offered 121 tools and it took a total of 7 hours and 18 minutes to activate them all, averaging 79 minutes per platform.

Time Spent		Number of Tools
Google Search Engine	15m	2
Facebook	3h 25m	33
Instagram	2h 31m	28
TikTok	2h 12m	25
Snapchat	15m	8
YouTube (main app)	46m	14
YouTube Kids	56m	8
Whatsapp	13m	3
Total	**7h 18m**	**121**
Average	**79m**	**15**

Our researcher classified the different tools according to difficulty of use. While the majority were very easy or easy to use, more than 20% required work, or were difficult.

Difficulty ratings					
	Very easy	Easy	Requires work	Difficult	Very difficult
Google Search Engine	50%	50%			
Facebook	29%	55%	13%		
Instagram	58%	23%	15%	4%	
TikTok	57%	29%	14%		
Snapchat	80%	20%			
YouTube (main app)	69%	31%			
YouTube Kids	50%	38%	13%		
Whatsapp		67%	33%		
Average	**56%**	**39%**	**17.60%**	**4.00%**	

To complicate matters further, tools change all the time. This shouldn't come as a surprise: apps are constantly changing, both in their functionalities and their shifting user base. On the platforms we considered, 29% of tools had changed in the last year; 13.75% in the last six months; 10.5% in the last month. On Instagram, more than one-third of user empowerment tools had changed in the last six months. Not only is there an enormous number of tools, but keeping up to date with exactly what they offer is a significant task. This is most definitely not a case of flicking a switch and then forgetting all about it.

Frequency of change						
	Monthly	6 months	Yearly	2 years	3 years	Stagnant
Google Search Engine			50%		50%	
Facebook		3%	12%	24%	9%	48%
Instagram	7%	7%	21%	21%	14%	32%
TikTok		12%	44%	24%		20%
Snapchat					17%	83%
YouTube (main app)				21%	7%	71%
YouTube Kids	14%		14%		14%	71%
Whatsapp		33%	33%	33%		
Average	**10.50%**	**13.75%**	**29.00%**	**24.60%**	**18.50%**	**54.17%**

> *'I didn't even think these apps had safety tools. I just rely on the safety settings within iPhone.'*

From our conversations with tech companies, we understand that many new tools are in development in response to the Online Safety Bill – both parental supervisory tools to help protect under-18s, and in reaction to the so-called Triple Shield, according to which adult users must be able to curate their feeds to avoid seeing, for example, racism, threats of violence against women, or homophobia.

There is a real danger of users being overwhelmed and bewildered. Parents will have to find and activate tools on platforms that their children use but they themselves don't. Far from inducing a feeling of user control, this may lead to a sense of being oversaturated with options, especially if tools overlap, or appear to, and it's not clear whether they interact with each other or offer slightly different functionality.

To make matters worse, different platforms may use different words for the same functionality – so that, for example, ‘muting’ and ‘hiding’ are different words for the same thing.

> *‘I work in school IT so I find the tools quite straightforward but I know a lot of people don’t. There are all sorts of safety systems that all have to be set up differently, and can be different depending on if you or your child have an Android or an iPhone. I think controls need to become standardised to make them easier for non-tech-savvy people to use.’*

It is not always easy for users to see what is on offer to them on a particular platform. Tools are often located in different areas of a site: what is needed are simple, user-friendly dashboards. Our researcher reported that Tik Tok was the only service where users could readily review their previous reports and, crucially, what action was taken. It is possible to review previous reports on Facebook, for example, but our researcher missed them initially, reporting that they were not especially intuitive, located within ‘Help and Support’ and then ‘Support Inbox’.

> *‘I have no idea where to locate them.’*

With tool usage being such an unwieldy and, to many, off putting process, we are concerned that parents should not be blamed for failing to implement them. It should not be parents’ fault when harms arise on platforms if they have not fully activated all the tools.

How effective are tools?

The most comprehensive review of the use and effectiveness of online parental control tools was published in 2021 by Svetlana Smirnova, Sonia Livingstone and Mariya Stoilova of the London School of Economics (LSE)[3]. Having screened 1,736 academic studies and looked in detail at 61, they reported that different types of parents used parental controls in different ways. They found significant variables, and policymakers cannot assume that if tools are provided, they will always be used, or used by all parents in the same way.

Parenting styles and values

Parents who liked rules used parental supervisory tools more often. So did those who wanted more active involvement in their children's digital lives. Parents who valued privacy used them less.

Risk aversion

Takeup depended on how prominent parents thought the risks were and how vulnerable they judged their own child to be.

Child development and need

There was broad agreement that some degree of monitoring was acceptable when children were younger. But there was no clear consensus about when this should end. We would expect this to be the case because children are different, their circumstances are different and parenting styles are different.

Parental tech skills

Younger and more tech savvy parents ***and*** the least confident parents tended to greater use of tools. Parents who felt that there were features that weren't clear to them used them less. To maximise takeup, tools need to be clear and transparent so that parents don't avoid them simply because they don't understand what they do.

> *'I'm more confident about some things than others. I'm unsure about the newer things, such as TikTok.'*

[3] https://euconsent.eu/download/understanding-of-user-needs-and-problems-a-rapid-evidence-review-of-age-assurance-and-parental-controls/

How well do they work?

Of the studies that the LSE researchers looked at in detail, 17 had found that parental supervision tools were beneficial or protective. But reduced screen time was used as a metric of success by some, and the consensus now is that time spent on screen is largely irrelevant; it's what users are doing that matters. A child can derive much more benefit from three hours of talking to friends than from 15 minutes of toxic scrolling.

When looked at in detail, the effects were more ambivalent, sometimes involving small or very small benefits. They were also affected by other factors, such as a child's impulsivity. Some measurements were taken without taking account of other protective factors, such as media literacy education in schools, so it wasn't entirely clear what had caused the improvements.

Eight studies found negative outcomes from the use of parental control and monitoring tools. These included family conflict, the erosion of trust between parents and children, and reduced feelings of privacy and autonomy among young people.

Six studies showed that, by reducing time online, the tools reduced access to educational content, limited the development of digital skills, and denied young people social interaction.

Twelve studies showed no effect from the tools, either because they were easy to bypass or because parents felt they didn't provide them with what they expected.

> *'We use privacy settings. They're easy to use but it can be inconvenient to get around to checking them as often as we should. Also, my child can change the settings if they wanted to.'*

Parent Zone's own research with the Oxford Internet Institute found that high levels of monitoring and control paradoxically led to more risk-taking behaviour[4]. Paradoxically – but also predictably: prohibition rarely works, especially for teenagers.

> *'I know where the tools are and they're reasonably easy to use but I don't trust that they're robust enough to always keep my child safe. Regular parental checks are of course still needed.'*

[4] https://parentzone.org.uk/sites/default/files/2021-12/PZ_Building_Online_Resilience_2014.pdf

In short, the parental monitoring and control tools we've had up to this point have very much not been a silver bullet. They will have to get a lot better if they are to satisfy both parents' and children's needs for them. And it must always be remembered that they will be used in different kinds of families, in diverse circumstances, with different values and ideas about how best to inculcate them.

A bit more history

Tools have often promised much – they sound so plausible – only to fall foul of the complexities of implementation. David Cameron's much-touted 'whole home solution' promised to cut off harms at the top level, at the router, before they got into the home. This sounded fantastic but only worked for those who accessed their internet via a router. For those who were going online via mobile telephony, it didn't apply.

There are also lessons to be drawn from the implementation of the EU's ePrivacy Directive and the subsequent GDPR (the General Data Protection Requirement), which, in changing the definition of privacy in the original legislation, caused a huge spike in cookie notices. The requirement for platforms to obtain user consent for the collection and use of their data was meant to be a jumping-off point for users to understand what platforms were doing with their data.

Unfortunately, people with jobs, families, hobbies – lives, in short – rarely have the time to wade through pages of legalese. And even if you reject everything possible on principle, there is often a complete lack of clarity. Is 'decline optional cookies' the same as 'accept selection'? What does 'continue with my choices' mean if you haven't actually made any choices? Do you move the slider so it's no longer green if the other side says 'consent'? Dark nudges urge you to accept whether you want to or not.

It's easy to develop opt-out fatigue especially if, after declining cookies and getting back on the site, the content goes dark. (It's not supposed to; it does happen.) Research has found that some websites store consent before offering a choice, so that in reality there's no opting out[5]. Work by academics at Aarhus University, UCL and MIT showed that 88.2% of cookie pop-ups on the top 100,000 websites in the UK were not complying with the law[6].

[5] https://www-sop.inria.fr/members/Nataliia.Bielova/cookiebanners/

[6] https://arxiv.org/abs/2001.02479

European privacy legislation, including the GDPR, is well-intentioned but has ended up being as frustrating as it is helpful. It doesn't give the sense of control that was promised; people may reasonably feel that they're being asked to do a lot of work for no discernible benefit. It is a technological solution that seems to penalise the user for not being like technology, for not being pre-programmed to manage cookies.

What about reporting?

> *'I don't think the moderators act quickly enough.'*

Parents tell us that when they report undesirable activity on a child's account, they want a swift response. This is not usually what they get. Often, they are left in limbo.

They also want to know that whoever is dealing with reports has safeguarding training at least equivalent to that of a UK teacher. Say that a 13 year-old girl reports that someone has asked her repeatedly for inappropriate images online: who deals with that report? In what jurisdiction? What responsibilities will be placed on them to report the matter to the police, social workers, parents? In some cases it will be better to alert one or other of these, but not all. There are fine distinctions to be made, and sensitive decisions with far-reaching implications. Will platforms be equipped to make them?

If reports are made and not dealt with properly or a response takes weeks or months to come, parents will lose trust, not only in reporting tools but in parental support tools more widely – and trust is crucial for the system to work.

What are the implications for data collection?

The original aspiration for the Online Safety Bill, to make the internet more closely resemble democratic societies that have evolved over centuries to be bearably civil, has been largely abandoned. The Bill is now focused on hoping to keep children away from hate speech, misogyny, pro-anorexia content and so on, with tools made available to allow adult users to curate their feeds.

To keep children away from toxic content, platforms will have to collect sufficient data to know if a user is a child, and what sort of a child – because if 17-year-olds are only allowed the same access as 13-year-olds, they will quickly rebel. This won't only affect children, of course: it will be necessary to ascertain whether each and every user is a child.

The big question is what will happen to all this data. As the Digital Futures Commission recently pointed out, data on children gleaned from EdTech in schools has been withheld from public bodies and civil society researchers because of data protection regulations on the one hand, while, on the other, it has been used by companies to pursue their commercial interests. The EdTech sector's data systems are opaque and its powerful players dwarf the capacity of a school to negotiate, 'or even to grasp the scale of their operations.'[7]

Once parents are made aware of the way data is extracted from their children, they become much more concerned about privacy. They expect to be asked for their consent. According to research by Rosalind Edwards, Val Gillies and Sarah Gorin, 81% of parents think that they should be informed as to how their child's data is being used[8]. It seems unlikely that this information will be offered up by data harvesting companies voluntarily.

We are still not clear what data exactly will have to be collected from children (or indeed adults), nor how it will be stored, nor whether it will be open to commercial exploitation. Recent history suggests that even if there are rules, ways may be found to break them.

It would be very unfortunate if in seeking to protect children from online harms, we bind them ever more tightly into the data-extractive surveillance economy, commercialising all their learning, socialising and play.

[7] https://educationdatafutures.digitalfuturescommission.org.uk/essays/introduction/problem-potential-childrens-education-data

[8] https://educationdatafutures.digitalfuturescommission.org.uk/essays/the-trouble-with-data/do-parents-trust-data-about-their-family

Conclusion

Parent Zone accepts that there is a role for technological solutions in mitigating online harms. But tools cannot do all the work, because so many other factors are in play – parental styles, media literacy and technological confidence, different levels of vulnerability and, crucially, trust. (Black parents and single parents are more concerned about the sharing of their children's data by public bodies, for instance.[9])

A slew of tools is already available. More are in development to meet new regulatory requirements. It is already possible to install parental controls at the level of the router, the device (including a console), platform, browser and app. The terminology used by all of these may well be different. They don't work to the same protocols. And they require vigilance if they are not to lead to a false sense of security: if parental controls are not available on a particular device, a parent might download Family Link, the Google product that helps parents manage their children's online activity. But this will only work with devices that are linked and compatible with the app. If the child picks up an unsupervised device or uses a different web browser, it won't apply.

Tools are not a fail-safe solution. They need to be easy to use, transparent, and to be seen to work. They need to earn users' trust, not least that of children. Children should not be incentivised to circumvent them. To achieve all that, tools need to respect children's rights to information, education, play and participation. They also need not to be seen as a method of harvesting children's and adults' data for commercial gain.

All of this will best be ensured by designing tools in conjunction with children and parents, transparently. Policymakers must also be aware that tools will be used in a complex ecosystem of family and social life, by families in very different circumstances, and with differing levels of media literacy. There is a tendency to talk as if tools override all context. But quite the opposite is true. Their success depends on the context in which they are deployed. We cannot afford to lose sight of that context – and in particular, of media literacy – nor to stop addressing it.

[9] https://educationdatafutures.digitalfuturescommission.org.uk/essays/the-trouble-with-data/do-parents-trust-data-about-their-family

Recommendations

- Tools should be designed with users, including children
- They should be easy to understand. They should not require users to read pages of legalese or dense terms and conditions.
- Ofcom should set a common language and protocols so that users understand what they are getting and don't have to learn new systems for each level of protection and each different platform.
- Ofcom should require tech companies to list all the available tools on their platform in plain English in a single, prominent, easily accessible place.
- Tools should be streamlined so that they don't overlap and duplicate. Each site should have as few tools as possible consistent with offering users the protection they require.
- Tools to keep children out of the adult internet should respect their growing capacity and need for autonomy up to the age of 18, and their rights to participate and to education, play and information.
- Ofcom should require platforms and services to respond to reports in a timely and proportionate manner.
- Ofcom should require platforms to put in place safeguarding measures for when children report issues that would trigger a safeguarding response offline.
- Platforms should allow users easily to review their past reports and the action that was taken.
- Tools should not be used to 'responsibilise' parents when harms arise.
- Platforms should prioritise minimal data collection.
- Platforms should not be allowed to sell children's data.
- Platforms should not be permitted to use opaque 'proprietary' systems and 'business confidentiality' to conceal what they are doing with children's data.
- There should be a public register of data collected on children, so children can see what has happened to their data.