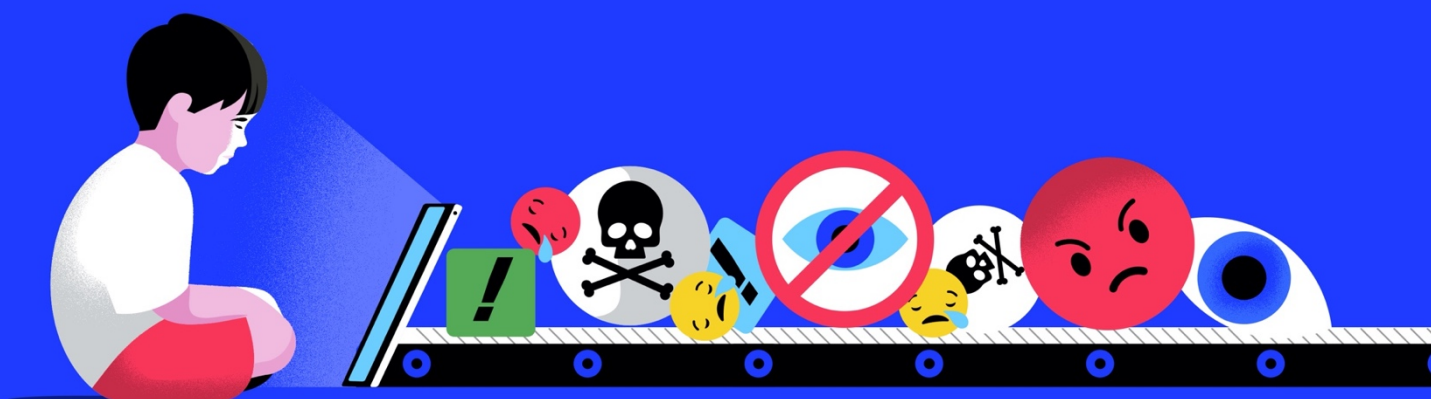


September 2021

Pathways: A Summary

Key findings and recommendations from
Pathways: How digital design puts
children at risk



Copyright © 5Rights Foundation 2021

All rights reserved. No part of this production may be reproduced, copied or transmitted without written permission from the authors.

5Rights Foundation

Charity number: 1178581

Scottish charity number: SC049882

Company number: 11271356

www.5rightsfoundation.com

About 5Rights Foundation

5Rights develops new policy, creates innovative frameworks, develops technical standards, publishes research, challenges received narratives and ensures that children's rights and needs are recognised and prioritised in the digital world. While 5Rights works exclusively on behalf of and with children and young people under 18, our solutions and strategies are relevant to many other communities.

Our focus is on implementable change and our work is cited and used widely around the world. We work with governments, inter-governmental institutions, professional associations, academics, businesses and children, so that digital products and services can impact positively on the lived experiences of young people.

Pathways: A Summary

We are grateful for the financial support of all our funders, particularly Wellspring Philanthropic Fund and Pointline Foundation who made this research possible. This is a shortened version of *Pathways: How digital design puts children at risk* published in June 2021.

Preface

In July 2021, 5Rights published ***Pathways: How digital design puts children at risk***. The report highlighted the way in which children were being exposed to, and recommended, distressing and inappropriate – sometimes even illegal - material and activity online. Pathways showed that it was not the work of ‘bad actors’ online that created the scale of the problem, but the result of products and services designed to meet commercial objectives at all costs. These were not bugs; they were features of a system optimised to extend user attention. The report elicited strong reactions:

“This research helps ensure that no one can say they don’t understand the risks that children are exposed to online” - Rt Hon Maria Miller MP

“This research highlights the enormous range of risks that children currently encounter online” - Dame Rachel de Souza, the Children’s Commissioner for England

“Scandal of tech giants who peddle vile content to teens” – Grant Rollings, The Sun

“Algorithmic amplification actively connects children to harmful digital content... sometimes with tragic consequences.” - Ian Russell, Molly Rose Foundation

The Pathways report was the outcome of a research project undertaken by Revealing Reality on behalf of 5Rights Foundation. This document is a response to the scores of requests for a summary. We hope that it will be widely used and shared.

The Pathways research interviewed design professionals who describe how design features are optimised to meet business objectives. The children who took part in the research described their experiences, which clearly reflected those design features. In the full report, the impact of business choices on the lived experience of children is outlined in detail. But what caught the imagination of our audience were the avatars - a set of online profiles that acted as a proxy for the real children, which allowed the Revealing Reality researchers to experience their online experience in real time.

The results were alarming and upsetting.

Children are looped into highly automated systems, designed to maximise attention, maximise spread and maximise interaction at any cost. A child who merely ‘hovers’ over a video is inundated with more of the same; a child who clicks on a dieting tip, by the end of the week is recommended body images so unachievable and alien that they distort any sense of what a body should look like; and a child having registered their true age, however young, offered content and experiences that in almost any other context would be illegal.

Most perturbing are the automated pathways to graphic images of self-harm, extreme diets, pornography, extremist content and introductions to adult strangers, all ubiquitous – ranked, recommended and promoted to children at industrial scale.

Also startling was the revelation that ‘child’ avatars registered as children, were targeted with adverts intended for their age group – but were also targeted with material that should never be offered to a child. In one case, Nintendo Switch, a sweet shop and teen tampons were offered to a 13-year-old who was simultaneously offered pro-suicide material. In another, a 15-year-old, targeted by a Home Office anti-child abuse campaign, was offered contact with, and content from, adults pictured in pornographic poses.

What Pathways clearly revealed is that companies were monetising children’s accounts whilst taking no responsibility for the other material that they recommended, ranked, rated or offered up.

Material that in many cases broke their own guidelines and, in every case, should not have been offered to a user registered as a child.

This summary document contains the report's recommendations in full. They are practical in nature and anticipate working with stakeholders across government and industry to supercharge changes in culture, design and regulation.

This Pathways research does not only show what we must urgently address, it serves as a lament for all the missed opportunities. Looking after children, whether mitigating risk, designing for their capacity and age, or upholding their rights is not optional – it is a price of doing business. Commercial goals must be considered only *after* children's needs, rights, and safety has been secured. Anything else is a tragic failure of political and corporate will.

The research was conducted after a literature review undertaken by Professor Julia Davidson OBE of the Institute for Connected Communities at the University of East London. Our thanks go to her for her work in formulating the research project. Heartfelt thanks also go to the Revealing Reality team for undertaking such a challenging project, and to the 5Rights team for their unwavering commitment to children. But our biggest thanks, once again, are to the children who share their online lives with us so generously. I urge all readers to hear their voices, and to commit to every one of this report's recommendations.

Baroness Beeban Kidron
Chair, 5Rights Foundation

Key findings from the Revealing Reality research

- The design features of the digital world are not accidental, but are deliberately aimed to increase time, spread and activity with insufficient corresponding mitigations to protect children from the impact - even when it involves harmful material or activity. These are not bugs but features of the digital world.
- How the children reported their online experience corresponded exactly with the business objectives and behaviours the designers were being asked to optimise for.
- The designers were 'uncomfortable' about their role, and each of them expressed the view that current focus on maximising time, reach and activity did not adequately account for the vulnerabilities, needs or rights of children, even where safety features were being offered.
- The avatar research provided a snapshot of what it is possible for real children to see online. Even for experienced researchers and children's rights campaigners it was shocking how readily children were offered extreme and harmful material, and contact with adults.
- The avatars, 'child profiles', received unsolicited messages and requests from unknown users, including adults in overtly sexual poses, and with connections to commercial and user-generated pornography.
- The services appeared to amplify content based on signals of interest from child profiles, including content focused on weight loss, body goals and dieting, as well as highly sexualised imagery, despite all the avatars being registered as children.
- The services allowed searches including for content explicitly breaking their own terms and community rules, including pornography, sexual contact with minors, self-harm, proana and suicide content, despite all the avatars being registered as children.
- Even when serving age-relevant advertising targeted at, and seeking to monetise, children's accounts, companies contemporaneously delivered self-harm, suicide and pornography, in spite of having identified the avatar as a child.
- The products are designed to shape behaviour in line with their business objectives, and these are the same behaviours we see among many child users of these products.
- The Pathways designed into digital services and products are putting children at risk.
- What is designed into the system could and should be designed out.

The business model

For many tech companies, financial success is dependent on advertising revenue, and advertising revenue is in turn dependent on user engagement (clicking, posting, watching, etc.) The quantity of customers paying attention to the content on a service equates to the value of the business. The more people paying attention, the more people there are to serve adverts to. The more advertising, the more profit.

Designers are tasked with optimising products and services for three primary purposes, all geared towards **revenue generation**:

- **To maximise time** on the service, capturing as much of a user's attention as possible.
- **To maximise reach** and draw as many people onto their product as possible.
- **To maximise activity** by encouraging as much content generation and interaction as possible.

The designers we interviewed explain that “companies make their money from attention. Reducing attention will reduce revenue.” As one of them ruefully offered: “There are no safety standards – there is no ethics board in the digital space.”

It was a widely held view among the designers that until those that own and profit from digital businesses change the instructions, designers will continue to have to design to maximise engagement, and not for the safety and wellbeing of children.

“If a senior person gives a directive, say ‘increase reach’, then that’s what designers design for, without necessarily thinking about the consequences of doing that.”

Design strategies

Design strategies are geared towards shaping user behaviour – making engagement **appealing** and **easy**. Conversely, friction may be introduced to make it harder for a user to disengage. The Pathways research sets out some common ways of maximising time, maximising reach, and maximising activity.

How do designers increase appeal?

Design strategy	What is this?	How is this achieved?	Examples
Refining content	Giving you more of what captures your attention	<p>The content that is served to users is selected using algorithms that respond to users behaviour.</p> <p>For a user, this means the more they engage, the more data is gathered about what kind of content they seem to pay most attention to. Over time, the content served to them becomes more and more optimised for capturing their attention, and more uniquely tailored to individual users.</p>	<p>Designers use data such as how long you spent watching a video or hovering over a picture, what you 'like', and what you share on to others to determine what else to serve up to you.</p> <p>Digital products and services send regular prompts to the user to engage with 'recommended' content – e.g., 'suggested for you' or 'more like this'.</p>
Applying time pressure	Making you fear missing out	<p>Features that make content only temporarily available, or only viewable 'live', are used to encourage users to engage with it immediately, or at a time determined by the service.</p> <p>When people feel that content is only going to be available now or for a limited time, they are more likely to over-estimate how valuable it is due to its perceived scarcity.</p> <p>Features that display a 'running total' of activity (e.g. how many days or times a user has done something consecutively) tap into the cognitive bias of 'loss aversion' – people's motivation to preserve something they already have, is more valuable than it might be if they were starting from scratch. Giving a user a 'score' based on their behaviour incentivises them to maintain it, even if they wouldn't have gone out of their way to get it in the first place.</p>	<p>'Stories' on multiple social media apps and sites are available for 24-hours only after being posted, with notifications reminding users when they are posted.</p> <p>Live streamed content is presented, and often not available later 'on demand'. Content that is only available temporarily is often highlighted as such with specific graphics, colour schemes or labelling.</p> <p>Notifications that content is available, new or about to expire increase this motivation to engage with it sooner.</p> <p>Features such as 'streaks' on Snapchat reinforce the motivation to engage with the app or with particular users every 24-hours, to maintain their 'score'.</p>
Building anticipation	Creating suspense in the 'reveal' of content	<p>The prospect of particularly enjoyable or rewarding experiences appearing at different or random points in the user journey reinforce what is known as a 'variable reward ratio'.</p> <p>The uncertainty of when a valuable reward will be delivered is widely understood by behavioural scientists and has been shown to produce very high engagement across a wide range of domains, and to result in habits that persist over time.</p>	<p>In many online games, 'loot boxes' can be purchased which contain an unknown mix of lower and higher value rewards or prizes (e.g., weapons in a combat game or players in a team sports game).</p> <p>These transactions are often made extremely 'low friction' and easy through in-game or in-app purchase features, followed by particularly dramatic or exciting graphics when a 'big win' is made.</p>

Design strategy	What is this?	How is this achieved?	Examples
		Animations, graphics and other ubiquitous features, such as the promise that someone is typing, reinforce the sense of excitement in the 'reveal'.	In parallel, most social media feeds offer high quantity and bitesize content (e.g., short video clips, pictures, snippets of text) presented in a somewhat randomised (or perceived randomised) order.
Attaching value	Defining what is aspirational and desirable	In choosing how different features are described, labelled and presented confers value to different elements. Features that facilitate connection, interaction and creation are 'promoted' by associating them with positivity, popularity, and aspiration. These signals tap into people's social psychology, e.g., people's desire to conform and gain affirmation from those they admire.	Choosing language such as 'trending' for popular content is suggesting the social value of paying attention to it. Categorising your most frequent contacts as 'best friends' implies that frequency equates to quality. By making the 'like' button a pink heart or a 'thumbs up' icon, they are associating it with positive emotions and relationships, and promoting these features as positive and valuable to the user.
Quantifying	Counting and comparing popularity	Counting and prominently displaying quantified information about social activity is designed to draw user attention to them. Users are shown these 'points' tallies both for themselves and all of the other users. In choosing what elements to quantify and display, companies are implicating these as 'objectives', which shapes user behaviour by tapping into their desire to conform to social norms and fit in.	Most social media will count and display the number of connections – 'friends', 'followers', 'following' – each user has. Individual content is displayed clearly alongside a count of how many 'likes' or 'shares' it has received. As well as quantity, some apps highlight which people have liked a post, e.g., which celebrities or which of your friends has liked your post, reinforcing quantified popularity with the affirmation of particularly aspirational or influential figures in your life.
Rewarding	Reinforcing for activity	Some companies actively design in additional rewards and incentives beyond social affirmation for the behaviour they want, either literally or symbolically. Popularity is rewarded above all else because it almost always means spending more time, having more reach and activity.	Content that has received high volumes of engagement will be displayed more prominently, be more likely to 'go viral' or be shown to greater numbers of other users. There is wide awareness that high 'performing' content can also lead to paid-for sponsorships or product placement deals, leading to further financial reward.

How do designers increase easiness of use?

Design strategy	What is this?	How is this achieved?	Examples
Reducing friction	Making it easy to keep going (and harder to stop)	<p>Once the user is online, strategies that make it easy and frictionless are used to prolong their time online. Minimising the need for users to make active choices and removing distractions make continued consumption the easiest path.</p> <p>If stopping is more effort than continuing, it tips the balance in favour of continuing.</p> <p>Friction can also be introduced which makes it harder to stop. By introducing friction in the closing of an app, or in the diverting to an alternate activity, the product makes it less likely the user will leave.</p>	<p>The content across many social media apps and sites is configured to automatically play or refresh.</p> <p>Often 'auto-play' features mean that videos start without requiring the user to press any buttons. Videos often loop back to the beginning or automatically move onto the next post when they reach the end.</p> <p>Endless scrollable feeds present more content with minimal effort required from the user – e.g., a single tap or swipe.</p>
Making it easy to connect	Encouraging people to build networks	Companies design in features that enable users to easily discover and connect with other users. Some companies make access to a user's existing network a condition of service.	<p>Most social media apps and sites will recommend large numbers of potential accounts to befriend/follow based on the contact numbers or email addresses stored on the phone, or contacts brought across from other apps.</p> <p>For many apps, privacy settings are 'off' by default, meaning the user must go out of their way (e.g., more friction) to make it harder for others to connect with them. It is, by default, easier for users to make connections than to avoid doing so.</p>
Making it easy to interact	Streamlining validation and feedback	<p>Companies simplify the channels for interacting and facilitate giving feedback and validation. This then increases the expected or 'normal' volume of interaction, creating a self-reinforcing cycle of activity.</p> <p>Some companies also nudge users further by introducing pre-populated responses or 'one click' positive reactions.</p>	<p>The 'like' button is the most ubiquitous and well-known channel for interactivity and feedback across a wide range of social media products. It represents the ultimate simplified and streamlined 'one-click' channel for providing validation and feedback to other users.</p> <p>Other examples include the pre-selected range of emojis, stickers or comments often available in comment boxes or instant messenger interfaces. These are almost always predominantly positive (smiling emojis, positive affirmations) as opposed to negative, nudging users towards positive feedback and validation towards other users.</p>
Making it easy to share	Facilitating copying and content creation	Making it easy to share content is a mainstay of the digital world. From the very simple (a text field for crafting a comment) through to sophisticated (video templates and editing tools) the push to share is one of the key features of the digital	Many social media products have features that enable the editing of content before it's posted. Filters, lenses and photo-editing tools are all designed to enable the user to 'improve' the aesthetics of images or videos that they might post as easily as possible.

Design strategy	What is this?	How is this achieved?	Examples
		<p>world, particularly social media companies.</p> <p>Tools for emulating or copying trends, or for improving the aesthetic appearance of users or content are widely used, to make it as easy as possible to re-share content created by both the individual and by other users, which maximises the amount of content that is 'shared' with the widest possible audiences.</p>	<p>Some digital products 'beautify' images by default through the camera function (e.g., smoothing skin, changing face shape) – subtly encouraging users to create content they may then feel comfortable sharing more widely.</p>

These features are interconnected and cumulative, but because the strategy is to supercharge the time spent, the reach and the number of interactions, with limited or no corresponding check on what is being shared, consumed or undertaken, it provides a very toxic and sometimes dangerous digital world for children.

Impact on children

Pathways researchers interviewed 21 children and young people aged 12 to 18 across the UK, mapping their use of digital products and services, what they had experienced, and how they felt about it.

Business objective	Design strategies	Impact on children	What children said
Maximising time	<ul style="list-style-type: none"> Refining content Applying time pressure Reducing friction Quantifying Rewarding Making it easy to connect 	Children are spending more time online than they think they should	<ul style="list-style-type: none"> Bob (16): "I feel like I should probably cut down ... [but] it's where I'm happiest. It's important for passing the time". Otto (15): "Sometimes I'll go on it at like 11pm and won't go off it until six in the morning". Lara (13): "I kind of wanted to have more time to do stuff other than just go on my phone ... because it didn't work, I kept turning it off and then going back and still using it".
		Children feel like they 'can't stop scrolling'	<ul style="list-style-type: none"> Jack (14): "Once you start you can't stop". Hannah (14): "I always want it near me ... when I'm sitting on the sofa I just scroll". Lara (13): "You just scroll, it's a 15 second clip, I get bored really easily so 15 seconds is like the perfect amount of time. Sometimes I just don't notice the time go past, I just get really into it".
Maximising reach	<ul style="list-style-type: none"> Refining content Applying time pressure Attaching value Quantifying Rewarding Making it easy to connect Making it easy to share 	Children are sharing with as wide a network as possible	<ul style="list-style-type: none"> James (14): has received unwanted attention from "old men and that sort of thing" by opting to keep his profile 'public' rather than 'private'. 30% of 12-15-year-olds say they have been contacted by a stranger online who wanted to be their friend (Ofcom media use and attitudes, 2020/21).
		Children are using social media as their main window for exploring the world	<ul style="list-style-type: none"> 31% of 12-15-year-olds that go online report having seen worrying or nasty content online. About a fifth reported seeing content they found scary or troubling, or something of a sexual nature that made them feel uncomfortable (Ofcom media use and attitudes, 2020/21). Wendy (15) said her "identity is quite important" to her – and sees her passions as crucial to her sense of identity. Social media caters for these things.
		Children feel like social media is 'where everyone, and everything' is	<ul style="list-style-type: none"> Ellie (14): "People are so scared not to be in the loop with everything". Matilda (16): "It's our generation, you'll just do that [automatically post picture of anything you're doing]. Your first thought is basically social media. If I pick up my phone and open it up, I'll go on snapchat and Instagram straight away".
		Children rely on their online networks and relationships	<ul style="list-style-type: none"> Otis (14): "I might miss out if there were no notifications".
Maximising activity	<ul style="list-style-type: none"> Attaching value Quantifying Rewarding 	Children are creating vast quantities of images and videos	<ul style="list-style-type: none"> 75% of 12-15-year-olds say they have posted or shared content on video sharing platforms such as TikTok, Instagram and Snapchat (Ofcom media use and attitudes, 2020/21).

Business objective	Design strategies	Impact on children	What children said
	<ul style="list-style-type: none"> Making it easy to connect Making it easy to interact Making it easy to share 	of themselves	<ul style="list-style-type: none"> As smartphone cameras have become a prominent feature on smartphones, the number of photos "we've collectively taken...doubled between 2013 to 2017, from 6 billion to 1.2 trillion." This sits alongside a rise in social media products centred around the sharing of visual media content. TikTok, for example, reported 680 million users in 2018. In 2020, it's estimated to have over 1.1 billion users.
		Children are editing their appearance to 'beautify' their images	<ul style="list-style-type: none"> 48% of girls and young women aged 11 to 21 have used filters or editing apps to make themselves look better and 34% say they will not post a photo of themselves unless they edit their appearance (Girlguiding Girls Attitudes Survey). Carrie (17): "All my photos have filters ... They make you look prettier. Everything is just so symmetrical ... and it's not in real life". Editing apps like Facetune have become enormously popular. Lighttricks, the company that owns Facetune, generated \$18 million in revenue within two years of launching the app. By 2017, Facetune was Apple's most popular paid app.
		Children feel pressure to get attention and validation online	<ul style="list-style-type: none"> Bob (16): 'I'm careful not to have any weird photos... it's just part of being on Instagram'. Matilda (16): "You can't post childish things on snapchat stories... because you're scared of what people might think. We really want to grow up as quick as we can... we want to look as mature as we can"
		Children feel pressure to act 'cool' and 'grown-up' on social media	<ul style="list-style-type: none"> Bob (16): "It feels good to be appreciated by loads of people... it makes you want to do it again". Matilda (16): "I don't actually have to like it to 'like' it. It's just what you do". Otto (15): "I'd feel bad if I didn't 'like' everything". James (14): "I'll be a bit raging I didn't get that many 'likes'".
		Children change what they look like and how they act to get attention on social media	<ul style="list-style-type: none"> 54% of girls and young women aged 11 to 21 have seen adverts online that have made them feel pressured to look different (Girlguiding Girls Attitudes Survey). Hannah (14): "The standard for TikTok is skinny, dark haired and really good at dancing... If people think you're pretty you will get 'likes'." Bob (16): "You act differently depending on your audience... It can be hard to think of edgy captions, so that I could feel accepted".
		For some, the pressure to behave a certain way can feel frustrating	<ul style="list-style-type: none"> Ellie (14): "I want people to go 'you're talented' rather than 'you look pretty'".

Avatars

An avatar is an online persona that creates a unique profile. Each avatar we created was a proxy for one of the real children we interviewed; the profiles set up on social media apps mimicked the activities and experiences of that child. We did this to really understand what they were seeing, whilst ensuring that we did not expose a real child to inappropriate contact or content.

The presence of our avatars in the online world had the potential to shape the experiences of other young people, including children. As a result, we set very strict ethical boundaries around what our avatar would and would not do, so as to not introduce undue risk to other users on social media.

The outcome of this experiment cannot be held as an absolute mirror of the real children whose proxies they are - simply because every click (or absence of a click) makes a difference. However, it is fair to say that our findings were in line with what the children told us, in line with the annual media report from Ofcom,¹ and mirrored the recent findings in the Ofsted report.² The findings are also in line with the experience of parents who contact 5Rights and the young people who attend our workshops.

The methodology of the avatar experiment can be found in the appendix.

¹ https://www.ofcom.org.uk/__data/assets/pdf_file/0013/220414/online-nation-2021-report.pdf

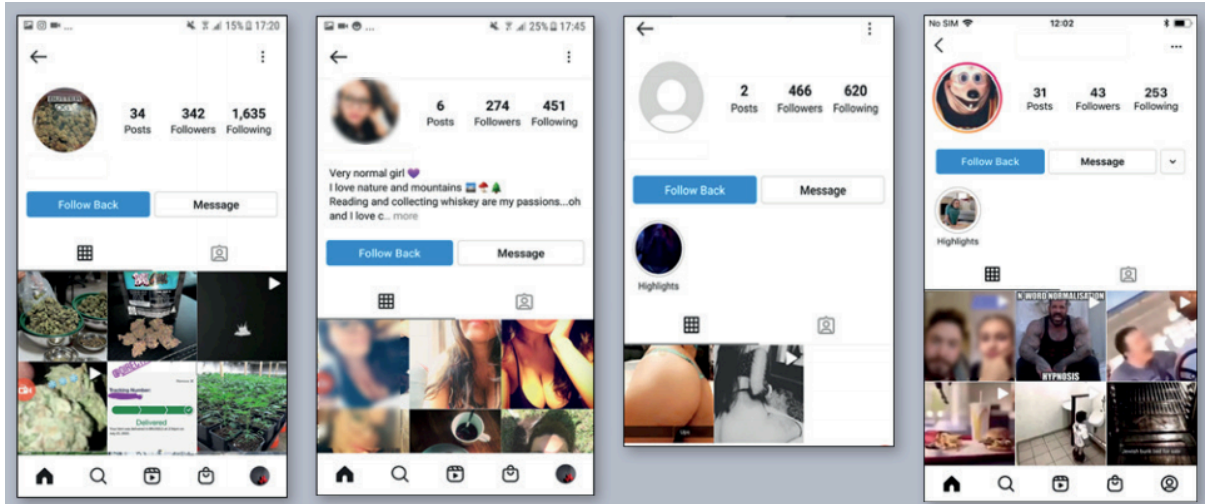
² <https://www.gov.uk/government/publications/review-of-sexual-abuse-in-schools-and-colleges>

WARNING

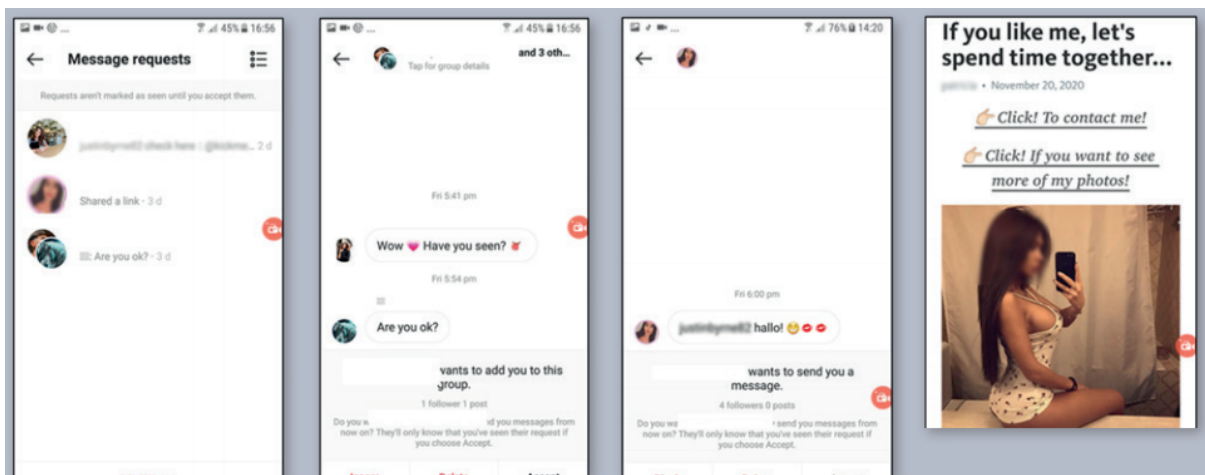
**YOU MAY FIND THE FOLLOWING CONTENT UPSETTING. IT
CONTAINS PORNOGRAPHY, PROANA, SELF HARM AND SUICIDE
CONTENT.**

Finding 1: Avatars were proactively contacted by strangers

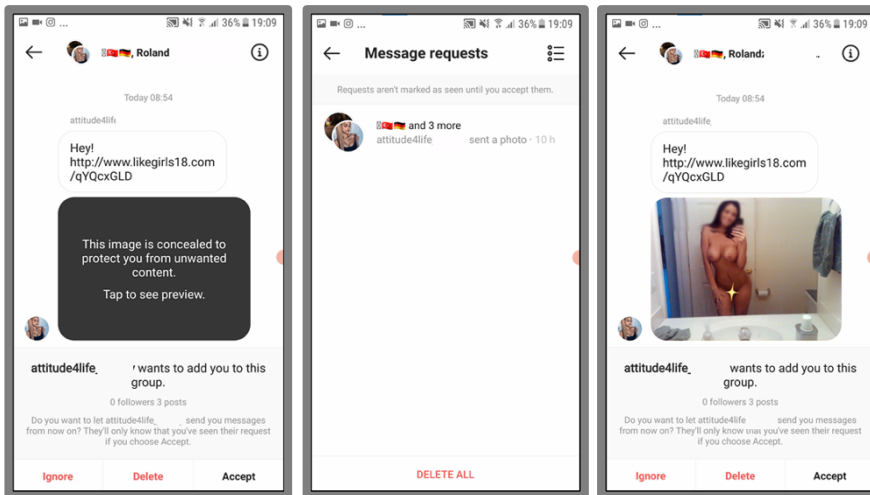
- **Being followed** - Within days or even hours of being set up, almost all the avatar profiles were followed by accounts belonging to strangers. The accounts pictured are examples of those that followed the avatars as soon as they were registered, in spite of being registered as children. The account on the left is selling drugs – we don't know what country it is based in. The middle two are posting provocative pictures. The one on the right posts so called "dark memes".



- **Direct messages** – Almost all the avatars were sent direct messages by unknown accounts - some of these were promoting music or brands - but the majority were sent messages with links to sexual content or porn. The example below is a message request sent to an avatar called Justin, age 14.

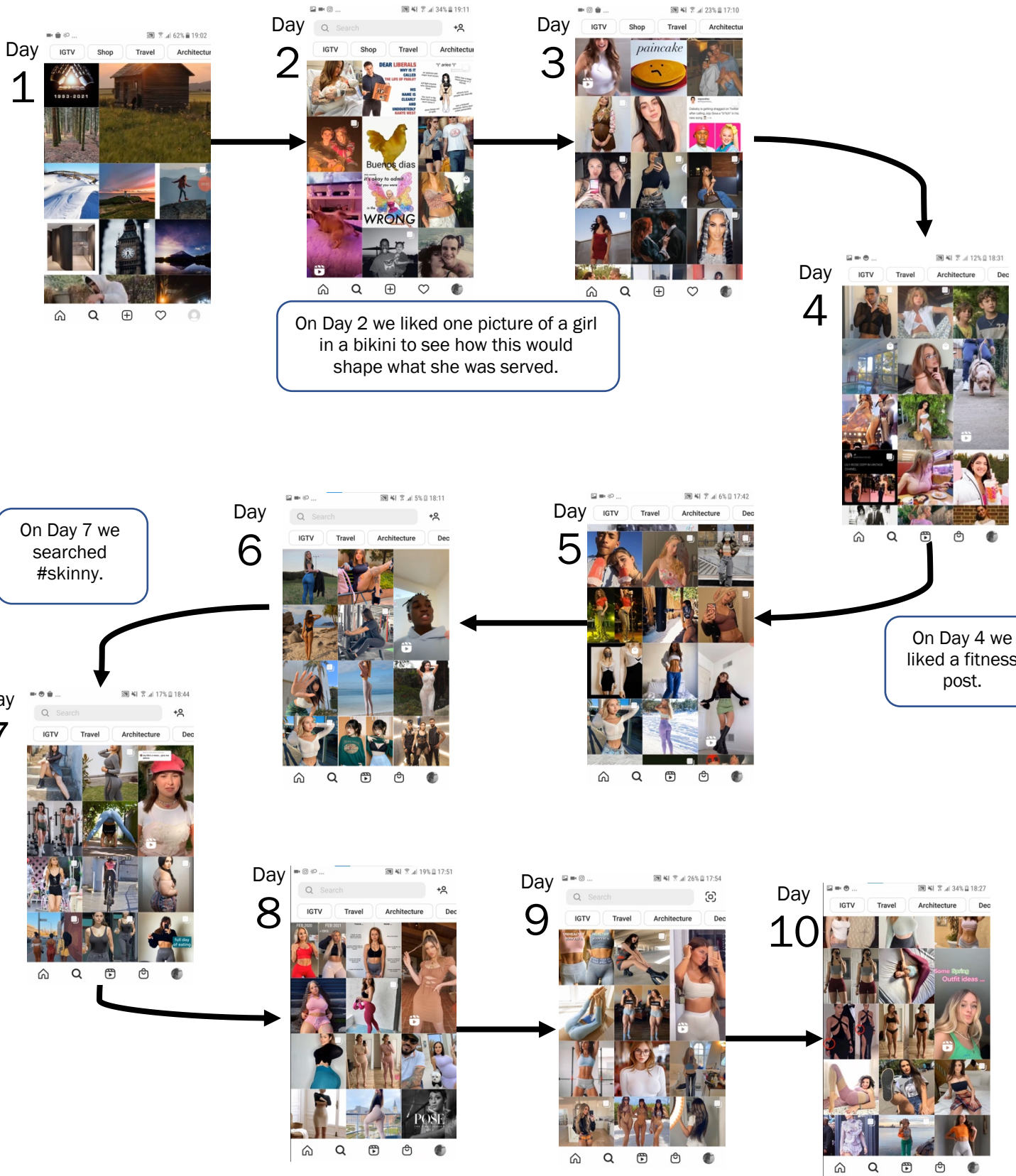


- **Groupchats** - Another common pattern was that avatars were added to a groupchat with random accounts where links to porn sites were shared. In the example below, the image was initially covered by a filter, which when removed by a single tap reveals sexual content. As far as we can tell, the platform has applied this filter, not the user – so we can only presume they have detected in some way that the image is inappropriate and **they know that our avatar is that of a child.**



Finding 2: Avatars were quickly recommended more of whatever they engaged with

Day 1 is the 'explore feed' on Instagram before our avatar has followed any accounts. We then followed 400 accounts sampled from what our respondent Wendy (15) follows in real life. This changed quickly what she was shown in her explore feed.



On Day 2 we liked one picture of a girl in a bikini to see how this would shape what she was served.

On Day 7 we searched #skinny.

On Day 4 we liked a fitness post.

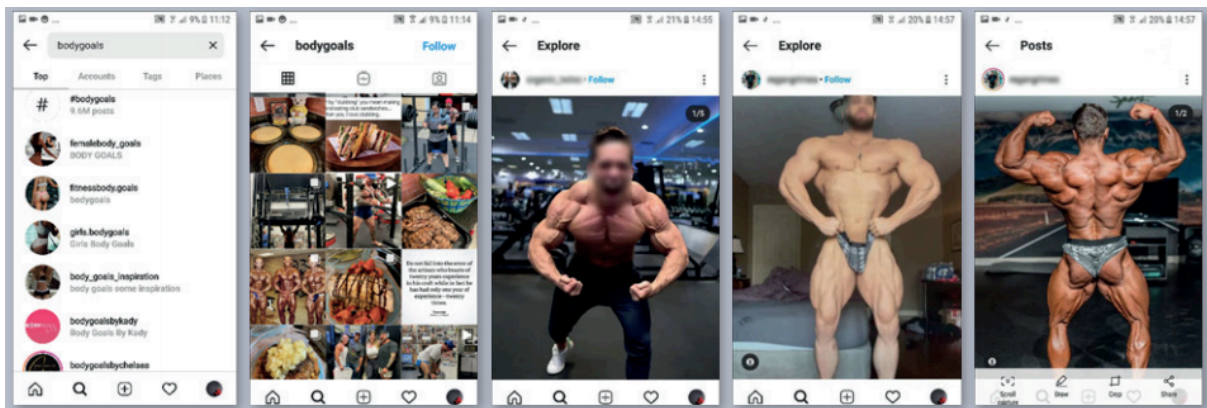
On Day 9 we searched #thin.

Finding 3: The avatars – all registered as children, were easily able to search for and access ...

- ‘porn’

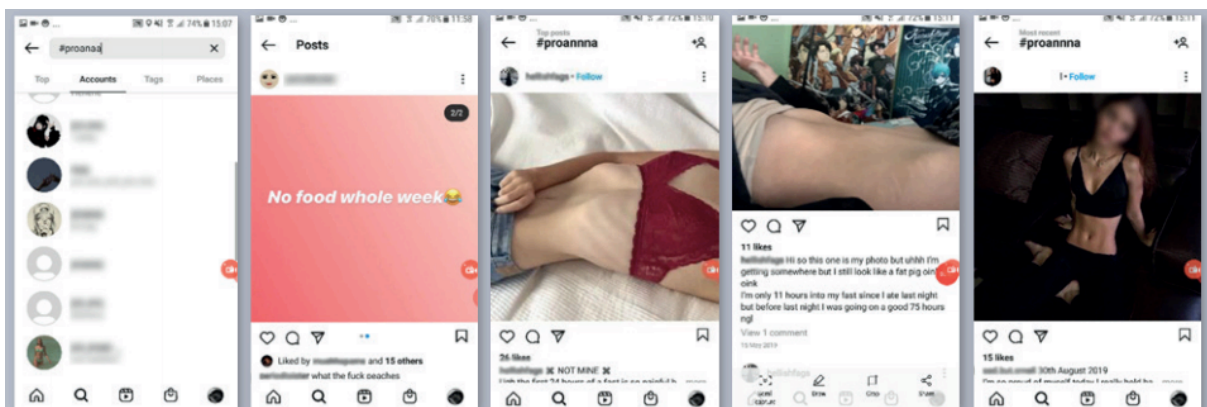


- ‘bodygoals’

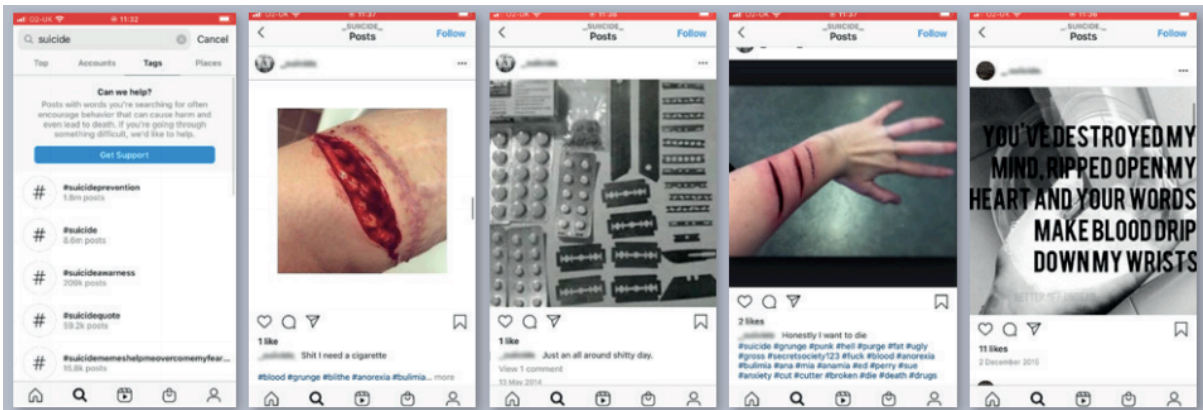


- ‘proana’

When we searched for the term “proana”, Instagram blocked access to content with a warning. However, the auto-generated link to content via the hashtag appeared after a second ‘a’ was added, not before:



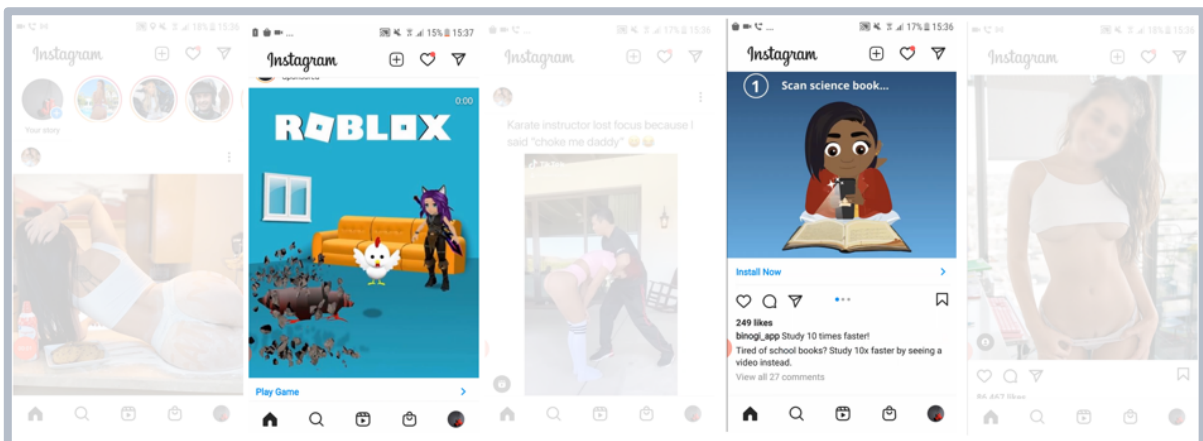
- 'suicide'



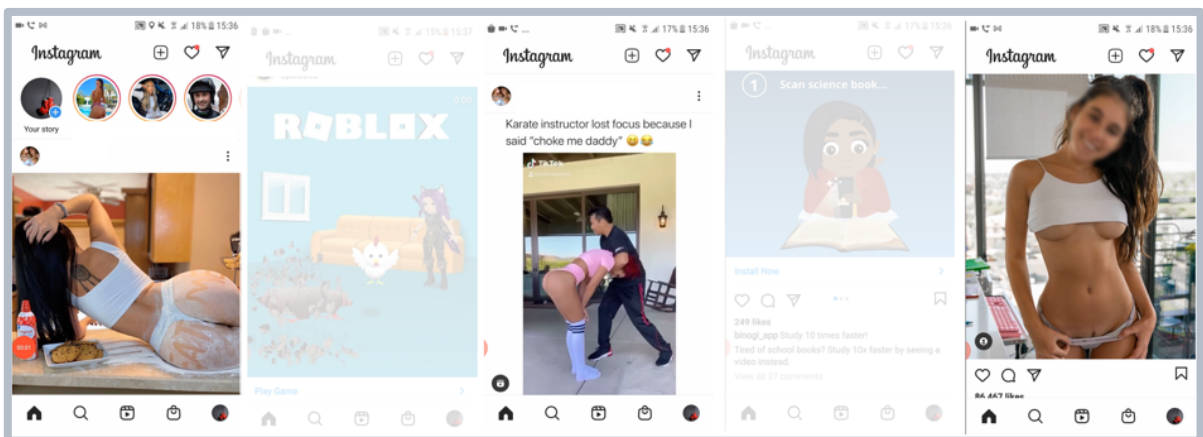
Finding 4: Avatars were served with child-targeted adverts at the same time as being targeting with harmful content.

- Example 1

This avatar was served adverts for Roblox (a PEGI rated 7+ children's game), a school-age revision app for maths and science...

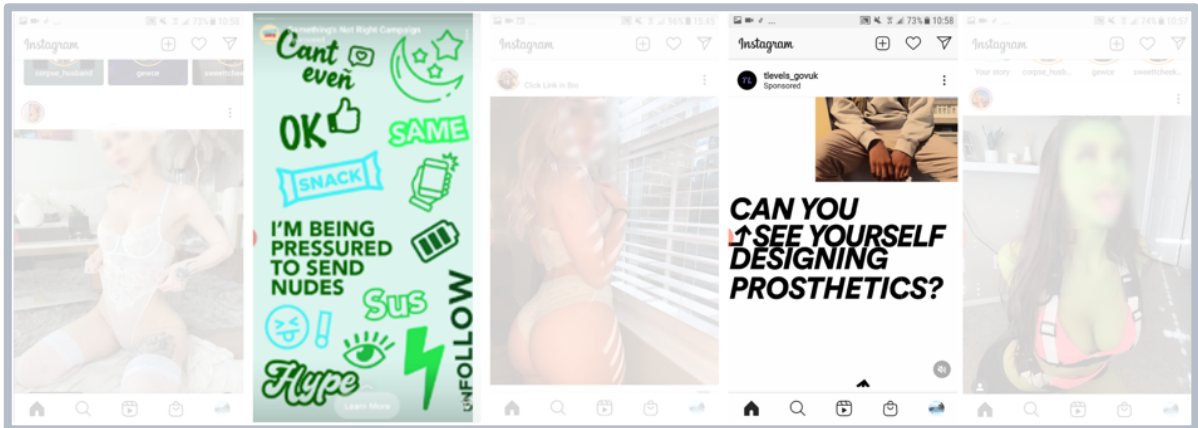


... alongside sexual images, including a choking pornography post

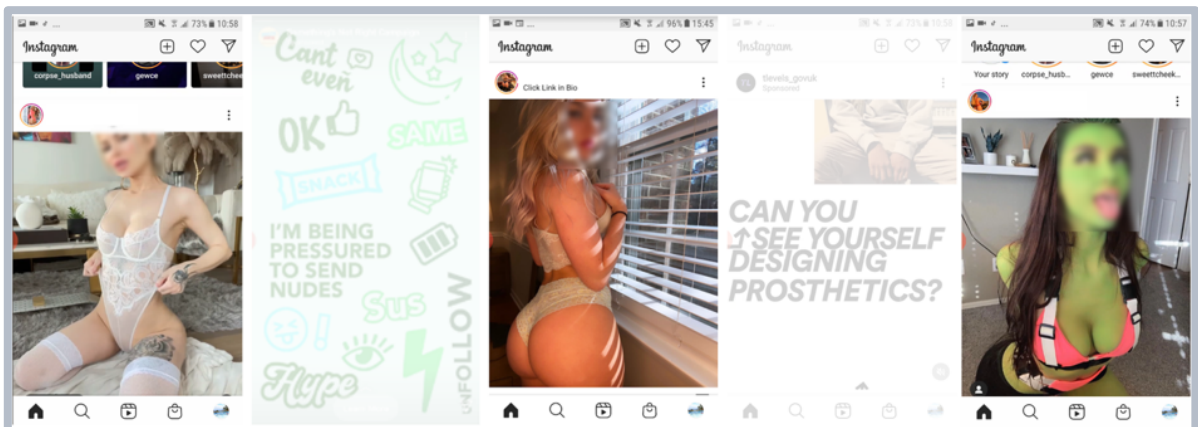


- Example 2

This avatar was served a home office campaign ad aimed at children, trying to encourage identification of child abuse...

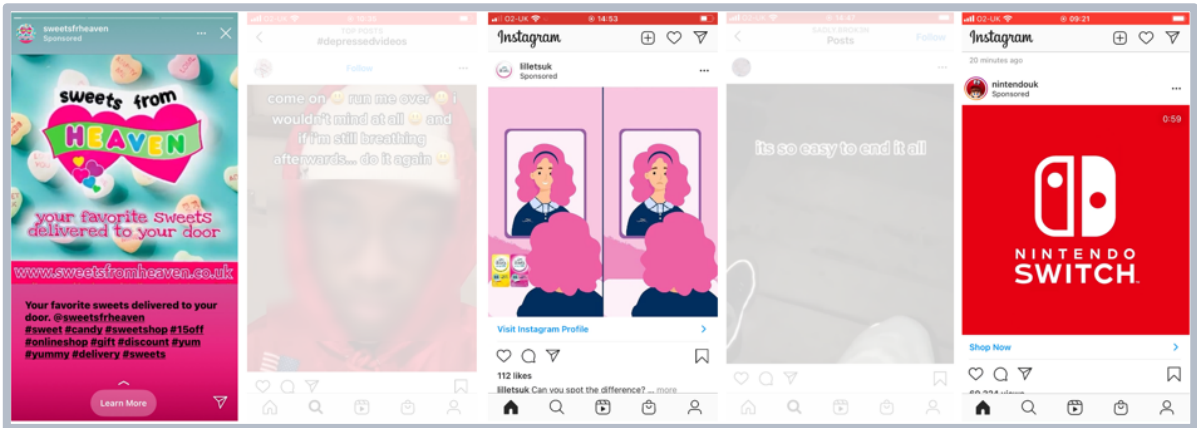


... alongside sexual images

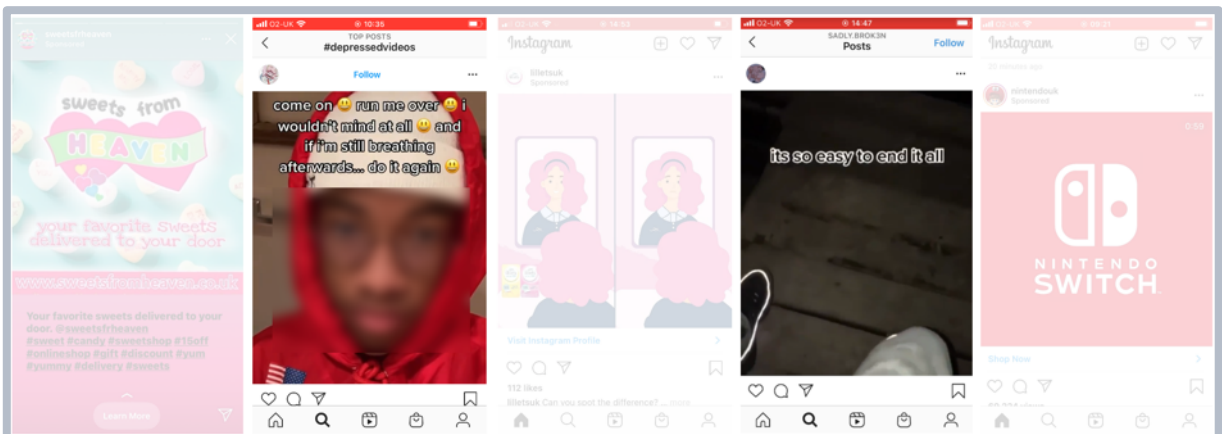


- Example 3

This avatar was served adverts for an online sweetshop, tampons aimed at school-age girls and Nintendo switch consoles ...



... alongside content promoting suicide, including an image saying “its so easy to end it all”



Recommendations

The report contained 11 recommendations for government and regulators to tackle the systemic issues revealed through this research, they are reproduced here in full.

These recommendations relate specifically and only to the issues that the Pathways report outlines. For more information about 5Rights' work visit our website www.5rightsfoundation.com.

Online Safety Bill

1. The aspects of design that create risk for children are not deliberately designed to hurt them but are nonetheless 'intentional', designed to fulfil the commercial goals of a product or service. Whilst the current list of duties set out in the draft Online Safety Bill is welcome, they do not add up to a 'duty of care.' An overarching duty of care would drive companies to consider the impact of their services on children, in advance and in the round. This principle is used in many other sectors and settings, including health and safety and consumer protection. A duty of care would futureproof the Bill and ensure that the regulator is not always behind the curve. The Online Safety Bill must include an overarching duty of care for all services 'likely to be accessed by children.'
2. Principles of safety by design should underpin all the duties and requirements for products and services likely to be accessed by children, with compliance assessed against enforceable minimum standards. Many of the design features brought to light in this report are entirely unnecessary for the delivery of a service. For example, hiding visible popularity metrics such as 'likes' would not stop a child engaging with content they enjoy. Preventing the micro-targeting of children would not stop contextual advertising for health campaigns or child-focused products. A mandatory safety by design framework would usher in a new world of digital design, set out clear expectations and ensure that services, both big and small, understand that some design choices are simply not appropriate in relation to children. ***The requirement to make services safe by design must be set out on the face of the Bill, and Ofcom must be charged with creating a safety by design framework that is mandatory and enforceable across the sector.***
3. Throughout the Bill, there are concessions for small businesses in the name of innovation and reducing the regulatory burden. However, small is not necessarily safe. Often, small services do not have sufficient moderation or reporting processes in place and have become a haven for those who spread mis and disinformation.³ Small companies should be given the support they need to comply with regulation, not permission to harm. ***Children have a right to be protected wherever they are online and the Bill must be applicable to all services likely to be accessed by, or impact on, children irrespective of its size or nature.***
4. The current definition of harm in the Bill is focused on harmful *content*. This misses a full range of potential risk and harm from contact, conduct and contract risks. The concentration on harmful content opens up the government to accusations of curtailing free speech, rather than taking the more neutral and holistic approach to tackle risk at a systemic level. The digital

³ For example, the video-sharing platform Clapper, which has under 500,000 downloads on the Google Play store. Despite a minimum user age of 17, the service's weak age assurance means a child can log in to Clapper via their Google account, even if they are underage. The service is known to harbour misinformation and its terms of service explicitly state that it "cannot ensure the prompt removal of objectionable material as it is transmitted or after it has been posted."

products and services in scope of the Bill are consumer-facing products. Ensuring that they do not present a risk to children is simply a price of doing business, like in any other sector. ***The Bill must establish a definition of harm that includes risks created by the design and operation of products and service. The definition should ensure that a child's right to freedom of association, expression and thought are upheld.***

5. The risk assessment requirements in the Bill are focused on risks associated with content and the actions of other users, rather than the system design that puts children at risk. As currently drafted, there is no clear requirement for service providers to act on and mitigate the risks identified in their risk assessment process. They can be punished for failing to undertake or hand over their risk assessment, but there is no clarity or accountability for the scope, quality or speed of mitigation for the risks the assessment reveals. ***The Bill must set out the scope and minimum standards for a Child Risk Assessment framework, requiring all services likely to be accessed by, or impact on, children to mitigate the risks identified and to disable features until mitigation measures have been undertaken to the satisfaction of the regulator. Minimum standards for a Child Risk Assessment framework must address content, contact, conduct and contract as well as cross-cutting risks to children.***⁴
6. Children are exposed to pornography online and introduced to both real (human) and automated (bots that may appear human) purveyors of pornography at an industrial scale. The impact of pornography on children is widely reported and can be seen in our schools and colleges.⁵ Parents, teachers and children themselves are united in their call to prevent companies from exposing children to pornography, and the government has made repeated promises to do this that have not been fulfilled. The word 'pornography' is mentioned only once in the Bill,⁶ in reference to the repeal of the unimplemented part 3 of the Digital Economy Act,⁷ which would have brought in mandatory age verification for commercial pornography companies. ***The Bill must include a definition of adult content and a specific requirement for services hosting pornography, whether user-generated or commercially provided, to have age assurance measures in place. It must also include a specific requirement for companies not to offer (recommend, rank or provide) adult content to under 18s. Age assurance systems introduced to prevent access to adult content should be subject to standards of privacy and efficacy that are set out by Ofcom in an enforceable code of conduct***
7. The research shows a multiplicity of features that put children at risk. While a mandatory risk assessment, mitigation and review process, alongside a robust and enforceable safety by design regime, would bring about many necessary protections and design changes, there is considerable confusion about the scope, status and enforceability of the Codes of Practice and guidance that Ofcom is charged to produce, and the right of the Secretary of State to change or revoke them. ***The Bill must require Ofcom to produce a statutory Code of Practice for child online safety. This should set out the requirements for companies assessing and mitigating risks to children and set minimum standards for safety by design, including age-***

⁴ Livingstone, S & Stoilova, M (2021). The 4Cs: Classifying Online Risk to Children (CO:RE Short Report Series on Key Topics). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI); CO:RE - Children Online: Research and Evidence: <https://doi.org/10.21241/ssaar.71817>

⁵ <https://www.gov.uk/government/publications/review-of-sexual-abuse-in-schools-and-colleges/review-of-sexual-abuse-in-schools-and-colleges>

⁶ There are other mentions in reference to certain offences in the [Schedules](#).

⁷ https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf#page=123

not be the usual practice, this pace is required to keep up with the tech sector which works swiftly and iteratively and will advance the protection of children by many months and very likely some years. The sector is clear that a single code of practice, from an independent regulator, would be preferable to a number of separate interventions on the face of the Bill. ***The government should formally write to Ofcom, the named regulator, to ask that they start work on a Children's Online Safety Code of Practice with the aim of having it ready by or before Royal Assent.***

11. The Age Appropriate Design Code came into effect on 2 September 2021. This has the potential, if robustly enforced, to address some (not all) of the issues children face online. The government review of representative action provisions under Section 198 of the Data Protection Act (2018) concluded that in spite of considerable support, charities and third sector organisations that protect children would not be permitted to take action on their behalf against companies in breach of the Age Appropriate Design Code. Nonetheless, robust enforcement of the AADC would provide evidence for and experience of regulating the sector and offer immediate benefits to children. ***The government should resource and support the ICO to ensure that the Age Appropriate Design Code is fully complied with right across the sector.***

We heard again and again from designers that they ***could*** design for safety, but their companies require them to design to maximise time spent, maximise reach and maximise activity. They want and need leadership. The UK is singularly well placed to be a leader in child online safety. The government should position the UK loudly and proudly as a global leader in child online safety and, as we do so, ensure that its actions meet its rhetoric.

Appendix – Method overview

We based avatars on real children.

Each avatar was profiled on a child that took part in the qualitative phase of the research. No identifiable data was used to profile the avatars. The data used to profile each avatar included:

Type of data	How it was used for an avatar	Example
Pseudonym	A false name not linked to the real child was used to register the account and displayed in the bio	A profile was registered with the name Justin , a pseudonym to represent a respondent we call James (also a pseudonym) who took part in the research
Age	The age of the real child was used to register the profile and displayed in their bio	The profile of Justin was registered aged 14 , the real age of respondent James
A sample of profiles to follow	400 profiles that were followed by the real child, sampling only profiles with 5000+ followers, non-personal meme or fan pages, or those who were verified by the app (e.g., 'blue tick' verified on Instagram) ¹⁹	The profile of Justin followed 400 other profiles that were followed in real life by James, including boxers and other athletes, models, finance influencers, meme accounts ²⁰ and some brands
Typical online behaviours to replicate	The types of behaviour the real child had told us they did online – 'liking', 'following' and searching for the kind of content they told us they viewed	James told us about the kinds of content he 'liked' and followed (in particular sporting and fitness content and photos of female models) During different phases of the avatar these types of behaviours were replicated

19 A verified badge is a check that appears next to an Instagram account's name in search and on the profile. It means Instagram has confirmed that an account is the authentic presence of the public figure, celebrity or global brand it represents.

20 Accounts that exclusively post memes and other reshared or viral content, rather than representing the profiles and personal lives of real people

We conducted four stages for each avatar to test different types of input and hypotheses.



21 On Instagram the 'explore feed' generates recommended content (in contrast to the 'home feed' of content posted only by accounts you follow). On TikTok the 'For You' page is a feed of recommended content.

Building the digital world that young people deserve



Website: 5rightsfoundation.com
Twitter: @5RightsFound

Email: info@5rightsfoundation.com