5RIGHTS FOUNDATION

Freedom of Expression Online

About 5Rights Foundation

5Rights Foundation develops new policy, creates innovative projects and challenges received narratives to ensure governments, the tech sector and society understand, recognise and prioritise children's needs and rights in the digital world. Our work is pragmatic and implementable, allowing us to work with governments, intergovernmental institutions, professional associations, academics, and young people across the globe to build the digital world that young people deserve.

A child or a young person is anyone under the age of 18, as defined by the UN Convention on the Rights of the Child. Rights language refers specifically to "children," however, children themselves often prefer to be called "young people." We use the terms children and young people interchangeably, but in either case it means a person under the age of 18, who is entitled to the privileges and protections set out in the UNCRC.

Executive Summary

- The right to freedom of expression is one of a set of rights afforded to children by the United Nations Convention on the Rights of the Child (UNCRC).
- When designed in the best interests of children and with regard for their developmental capacity, digital technologies can provide important ways for children to express their views and have their voices heard.
- Expression online, particularly for young people, is not only text-based 'speech' but increasingly self-generated, image-based content.
- Many young people do not feel they can express themselves freely in the digital world and vulnerable young people are disproportionately impacted by threats to freedom of expression online.
- Aggressive data collection and processing creates a digital footprint that has a 'silencing effect' on young people, which can curtail or limit what they say, do or post online.
- The business norm of retaining data denies children the opportunity to erase or grow out of what they have said and done online.
- The widespread use of profiling across digital services has a detrimental effect on both a child's freedom of thought and freedom of expression.
- Use of algorithms in recommendation systems can undermine a child's right to freedom of expression, to access information, their right to participation and their freedom of thought.

- Companies must take into account their role in spreading, promoting and recommending content – this is inseparable from the question of freedom of expression and the right to access information.
- A fundamental cultural shift is needed to stop companies from using automated decision-making systems that power seemingly benign recommendation features, until sufficient safeguards are engineered into the design of these systems to protect children's freedoms.
- Strengthening competition regulation of big tech companies will create an environment in which young users are able to make more meaningful choices, give them greater autonomy over how and in what way they choose to express themselves, and create a more diverse digital ecosystem.

Introduction

The digital world has created a seismic shift in the way people communicate, access information and share their views. With social media, anyone can broadcast their thoughts to millions of people in an instant – a freedom and power unimaginable to most thirty years ago. In this context, freedom of expression has taken on new meaning and significance. It has amplified marginal and previously unheard voices but given voice to those who seek to confuse, undermine or spread falsehoods. Illegal content such as hate speech, incitements of violence and child sexual abuse proliferate online, with tech companies failing to mitigate, and facilitating the spread of illegal material on their services. Equally, content and behaviour that may not be illegal but causes serious harm is spread widely, including disinformation, misogyny, trolling, and pro-suicide content, and is often left unchallenged, unreported, and unremoved.

Central to this picture are the business models and content monetising standards of the companies that offer opportunities for communication. Designed to promote and extend engagement, they supercharge the distribution of information largely based on automated popularity metrics and commercial interest. Promoting content against these criteria, rather than the veracity or source of the material, has allowed a torrent of mis and disinformation to flood the digital world, in particular, across social media. While the impact of this has proved challenging for adults and society generally, it places particular and additional burdens on children - a demographic that gets more of its information online, but lacks the maturity, life experience and financial resources to access trusted alternatives. And all at a time of life when both the faith in and the hurt from extreme views is developmentally normal.

5Rights notes this inquiry is taking place as the UK government begins drafting new legislation for the Online Safety Bill, and the Law Commission consults on proposals to reform communications offences. It is our hope that the Committee's inquiry will ensure that freedom of expression is put in balance with the government's stated ambition "to

make the UK the safest place to be online."[1] Additionally, that the inquiry will pay significant attention to the way in which views are promoted, ranked and spread on digital services. To this end, the Committee should also consider the right to freedom of thought - a necessary pre-condition of the right to freedom of expression – and the ways this fundamental freedom is threatened in the digital world.

5Rights Foundation speaks specifically on behalf of and is informed by the views of young people. Therefore, our comments reflect, and are restricted to, the experiences of young people under the age of 18. However, we recognise that many of our views and recommendations are relevant to other user groups and we welcome any efforts that government makes to make the digital world more equitable for all user groups, particularly the vulnerable.

[1] https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response#annex-a

Responses to consultation questions

i. Is freedom of expression under threat online? If so, how does this impact individuals differently, and why? Are there differences between exercising the freedom of expression online versus offline?

Is freedom of expression under threat online?

Yes.

1. The right to freedom of expression is one of a set of rights afforded to children by the United Nations Convention on the Rights of the Child (UNCRC). This also includes the right to leisure and play, the right to access information, the right to privacy, freedom of thought and the right to protection from violence.[2] The fundamental premise of these rights is that they are indivisible and of equal importance. A child's right to freedom of expression cannot therefore be considered in isolation from their other rights in the digital world, particularly their rights to privacy, freedom of thought and protection from undue influence. In the context of the digital world both industry and lawmakers have privileged freedom of expression at the expense of other rights which has created a distorted digital environment for young people.

2. The technology employed by digital products and services is not neutral. It promotes and popularises information using opaque criteria and for largely commercial purposes. This means that voices are not given equal weight and are algorithmically prioritised. Research shows that more extreme, more sexualised, more aggressive content travels further and that digital services promote this content to drive engagement. This 'intervention' undermines the fundamental concept of freedom of expression that gives each person an equal opportunity to contribute their voice. In the case of children, it creates a toxic environment in which many children are afraid to speak in case of extreme or aggressive responses.

The attention economy is based on the greatest rewards of attention being given to the loudest, sexiest, most opinionated, outrageous, bravest or tragic. The need for attention is problematic for children who do not yet know how to judge the veracity of what they are attending to, and who are vulnerable to making long-term decisions for themselves about their digital identity without understanding the commercial purposes of the digital environments they are

1. [2] The rights granted to children under the UNCRC are assigned a special status under international law. As stated by the European Court of Human Rights, this status as *lex specialis* requires that the European Convention and other international treaties are interpreted in a manner that gives effect to children's rights. This requirement extends to the Court's interpretation of the rights and obligations conferred under Article 10 ECHR, where it is noted that the best interests of the child must be given "paramount importance".

inhabiting, and without having sufficient access to the creative and participatory elements of the technology they are using. Testing the limits of sexuality or popularity is not new, but the environment in which things are shared, copied, commented on and amplified exponentially, is.[3]

5Rights' 'Digital Childhood' report

3. Data collected by digital services create a digital identity for a child. It is estimated that before the age of 18, more than 70,000 data points will be collected.[4] Children are increasingly aware that what they say or do online will form part of their digital identity and will be used to make decisions about them, including their access to education or work opportunities, their credit rating and the cost of services in the future such as insurance. This can have a 'silencing effect' that results in them curtailing or limiting what they say or do online, and therefore has a significant impact on their right to freedom of expression.

"My view on the digital world has changed and I am now more aware of what I'm agreeing to and how companies use my data." - *Young person aged 17, interviewed by 5Rights*

4. A child's understanding of the world is constantly changing and with it, the thoughts and feelings they use to express themselves. A child needs to test boundaries and try out new social interactions as they grow. **But the business norm of retaining all of a child's data – forever – and the way that data is added to an ever-increasing footprint means that children are denied the opportunity to leave behind or grow out of what they have said and done online**. Crucial for the Committee to consider is that a child's digital identity is shaped not only through the companies and activities with which they choose to engage, but also by those of their 'network', their locations, their gender and other markers. This has an enormous impact on their freedoms, including that of expression and association, since it is this 'other self' and the views of others in a child's network that is being acted upon.

"Personally, it's like when you're younger, you'll do things, but you'll look back on it and you'll regret it - and if you regret it that much you should be able to delete it and pretend it never happened. So if they wanted to delete it because they were younger, they should be able to because obviously... if they've changed, if they're embarrassed by it, or if they feel like they've improved on something, they should be able to get rid of the previous thing." - *Young person interviewed by 5Rights*

5. The digital world, and particularly social media, has made a child's inner world available for all to see. Broadcasting fleeting thoughts and livestreaming intimate

[3] https://5rightsfoundation.com/uploads/digital-childhood---final-report.pdf
[4] https://www.childrenscommissioner.gov.uk/wp-content/uploads/2018/11/cco-who-knows-what-about-me.pdf

moments are actively encouraged by digital services, which have normalised design features that nudge children into sharing with the world, often in real time, their interactions, mood and thoughts. Features such as 'likes' 'shares' and 'follows' exploit the need for social affirmation, which during adolescence - a time of increasing autonomy and identity exploration – has a significant impact on self-esteem. While services create experiences that can be exciting and fulfilling for young people, they also provoke anxiety from overexposure, mass social judgement and a lack of control over self-image and personal data, all of which are forms of expression in the digital world. In particular, the norms around image sharing have created a strong feeling that young people cannot share images of themselves without using filters and enhancement.[5] **It is important that the Committee consider images and not only text, particularly self-generated images, as forms of expression. Also, how algorithmically spread comments and messages impact on a child's ability to exercise their right to freedom of expression.**

Tweens talked of using filters for contouring and making your cheek bones more prominent, as well as making your skin look flawless. One child said that the changing of your appearance is not just for your own feeling of self-worth, but "to make people interested in them", illustrating that popularity online is to do with appearance rather than personality. "I wish I was wearing a filter right now," one girl said.[6]

6. Young people are more likely to share news and opinions in messaging apps due to concerns about the potential negative impacts to their reputation of sharing in more public channels. Concerns about visibility and the consequences of sharing to large and sometimes unknown audiences online also have a disproportionate impact on particular groups who feel vulnerable, for example. Nearly half of girls admit to holding back their opinion on social media for fear of being criticised and 13% of girls have stopped going on social media altogether to avoid negative responses.[7]

   Many young people do not feel they can express themselves freely in the digital world and that vulnerable young people are disproportionately impacted by threats to their freedom of expression online.

<u>Are there differences between exercising the freedom of expression online versus offline?</u>

7. The debate on freedom of expression has moved on from where it was even a few years ago, when many still considered the digital environment to be a sacred and separate space in which freedom of expression trumped all other rights and

[5] Over a third (34%) of girls and young women aged 11 to 21 will not post a photo of themselves unless they change aspects of their appearance, and this increases as girls get older. Source: Girl Guiding UK.
[6] https://theconversation.com/i-wish-i-was-wearing-a-filter-right-now-why-tweens-need-more-emotional-support-to-deal-with-social-media-149876
[7] https://plan-uk.org/act-for-girls/girls-rights-in-the-uk/reclaiming-the-internet-for-girls

protections. It is now generally accepted that technology is a seamless part of the fabric of a young person's existence, and that the binary of 'offline' and 'online' is no longer in keeping with the reality of everyday life. The digital world is not optional for young people. It is their gateway to education, information, entertainment, health services, and mediates their relationships and experiences.[8] If it is not optional then the rights, entitlements and protections children enjoy offline must apply online.

8. There remain stark differences between the regulation, legislation and enforcement of rights and protections online and offline. The sheer scale of large digital services presents a challenge for moderation systems to police individual pieces of content. Equally, the lack of industry regulation around 'terms of use' and gaps in legislation. Where legislation does exist there is poor enforcement, for example the very low prosecution rates for online hate crime.[9] The Committee could usefully make recommendations about resources and commitments to enforcement as well as seeking to plug the legislative gaps.

9. There is also a generational gap, where older people sometimes mistake a young person's 'facility' (two fast thumbs) for 'understanding' - having knowledge and agency in the context of digital world. In spite of their dependence on and facility to operate digital products and services, children remain resolutely at the bottom of the digital ladder.[10] It is important that the idea that they are safer or better at 'being online' than adults is not baked into our response to the digital world.

ii. How should good digital citizenship be promoted? How can education help?

"The older you get the more experience you have and you know what's right." - *Young person interviewed by 5Rights*

10. Teaching good digital citizenship should not be separate from teaching about the purposes and likely outcomes of digital use, with an emphasis on data literacy. 5rights has conducted a number of deliberative workshops with young people, during which it has been clear that knowledge of how the system works gives young people confidence, promotes behaviour change and encourages critical understanding. Comprehensive and age-appropriate education and training should be available for children of all ages, throughout their schooling. Parents, carers, children, teachers and frontline workers need high-quality information that promotes digital citizenship, data literacy and agency.

[8] https://www.unicef-irc.org/publications/pdf/idp_2016_01.pdf
[9] 1,605 online hate crimes were recorded in England and Wales in 2017/18, with the estimated rate of arrests ~10 per 100,000 people. (See: https://www.turing.ac.uk/sites/default/files/2019-11/online_abuse_prevalence_full_24.11.2019_-_formatted_0.pdf)
[10] https://hansard.parliament.uk/Lords/2014-11-20/debates/14112059000611/UNConventionOnTheRightsOfTheChildDigitalImpact

11. The current provisions do not go far enough In the statutory guidance for Relationships Education, Relationships and Sex Education (RSE) and Health Education, digital literacy makes up just 1 core module out of 8 for both primary and secondary school under the umbrella of 'Physical health and mental wellbeing' and 1 core module out of 5 in the vein of 'Relationships education.'8 [11] This does not constitute meaningful provision and is out of kilter with the impact of technology on young people's lives.

12. Education and resilience building in children must not be the answer to addressing the problems of the digital world. Products and services must be designed in ways that do not promote extremism and that balance the right of the originator's freedom of expression with the full gamut of children's rights. Children must not be penalised for behaving in ways that are facilitated or encouraged by the design of services, and service providers must take responsibility for helping children make informed decision and give meaningful consent when engaging with a service. Moreover, children need technology to be responsive to their needs and capacities at different stages of development. They should not be expected, particularly children in the youngest age groups, to adapt to the structures of technology developed with adults in mind.

13. **It is inappropriate to try to educate young people to operate in a world which systematically asks them to act beyond their maturity and puts them at risk, and it is dangerous to make them responsible for aspects of design over which they have no control.**

iii. Is online user-generated content covered adequately by existing law and, if so, is the law adequately enforced? Should 'lawful but harmful' online content also be regulated?

14. Currently user generated content is not adequately covered by existing law. In its proposals for the new Online Safety Act, the government has set out number or requirements that services will need to fulfil under the new duty of care to address harmful user-generated content. However, much of the language used in these proposals puts freedom of expression in contest with online protections.[12] Almost all measures to protect users online are qualified by a commitment to defending freedom of expression. The proposed regulatory framework gives woefully little attention to the design systems, risky features and business models of the tech companies that allow such harms to promulgate and in many cases are themselves responsible for spreading harm. While making a welcome set of proposals on behalf of children, the proposals do not fully recognise the principle of indivisible and interconnected rights, nor do they put sufficient protections on the face of the Bill.

[11] Relationships Education, Relationships and Sex Education (RSE) and Health Education, Department of Education, 2020.
[12] https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response#contents

15. If the proposals contained within the government's response come into force, the Online Safety Bill will introduce requirements for companies with the largest online presence and high-risk features to address both illegal and legal but harmful content on their services. Companies in scope must "take action to prevent user-generated content or activity on their services causing significant physical or psychological harm to individuals."[13] There will also be "a specific legal duty to have effective and accessible reporting and redress mechanisms… [to] cover harmful content and activity, infringement of rights (such as over-takedown), or broader concerns about a company's compliance with its regulatory duties."[14] 5Rights has serious concerns about the scope, the definition of harm and the proposals for secondary legislation and enforcement.

16. When considering risks to young people online, the emphasis frequently settles on the most extreme harms, such as grooming and child sexual abuse. In reality, mercifully few young people suffer acute harm, but many are victims of the cumulative effects of so-called 'lesser' harms, including those caused by lawful but harmful content - content whose overwhelming presence in the digital world is justified on the basis of freedom of expression.[15] Such content includes the promotion of eating disorders, online bullying and misinformation, including health misinformation. The social and financial cost of these harms has yet to be calculated, but it is clear that the burden on education, health, mental health, police, local (council) support services, as well as the individual, family and community costs are rising exponentially. **The Committee should consider if the freedom to post such material extends to companies promoting, sharing and or recommending it to children**.

iv. Should online platforms be under a legal duty to protect freedom of expression?

17. Tech companies have assumed traditional governmental functions in their ability to regulate online public spaces. There is agreement among both freedom of expression and child protection advocates that 'soft law' attempts have failed to encourage tech companies to exercise these powers in accordance with human rights law, such as through the adoption of codes of practice.[16] This has culminated in calls to consider whether it is now time to extend the human rights obligations they hold under international law to private companies.

18. The proposed legislation set out in the government's Online Harms response includes safeguards for freedom of expression and pluralism online, protecting people's rights to participate in society and engage in robust debate. While this is welcomed, service providers should have a legal duty to protect *all* rights held by

[13] https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response#annex-a
[14] Ibid
[15] https://www.riskyby.design/introduction
[16] see the United Nations Guiding Principles on Business Rights ("the Ruggie Principles"), 2011

children, and design and operate services in a way that considers their best interests as paramount. This will protect not only a child's freedom of expression, but their freedom of thought, their right to participation and their right to access information.

19. Companies should also have a legal duty to uphold the rules and standards they set out in their published terms. 5Rights research shows that 82% of British parents agreed that internet companies should be held accountable in law for how well they uphold their own community guidelines, terms and conditions, and privacy notices.[17] The UK's Children's Code includes a requirement for service providers to uphold their own terms, explaining that young people "should be able to expect the service to operate in the way that you say it will, and for you to do what you say you are going to do."[18] This should be reflected in the forthcoming Online Safety Bill, requiring regulated services to publish community standards and other terms that meet a set of minimum standards established by Ofcom. Published terms must be presented in ways that are truthful, easily understood and accessible to young people accessing the service, at the time or times when they are most likely to engage. Services must also be legally required to put in place clear processes to ensure that their service's community standards and other published terms are upheld, including what action will be taken if they are violated.

20. **Above all, any duty on a company must take into account their role in spreading, promoting and recommending content – this is inseparable from the question of freedom for the originator of content.** In a recent example, when QAnon was taken down by twitter, a senior member of the House of Commons lost 70 hostile followers overnight. The network effect is such that the female politician who the followers were trying to silence was enabled by Twitter. **We support any companies right to host someone's content, and even allow it to be found - but not their right to facilitate its spread.**

v. What model of legal liability for content is most appropriate for online platforms?

21. The digital world is no longer a communication tool, but rather impacts on every part of society from science to education and health, from entertainment to the justice system. Any models of legal liability are likely to be complex and interconnect with other areas of law, and should be subject to frequent review. 5Rights suggests that the following multi-level system of legal liability is needed to successfully regulate online intermediaries.

(a) Liability under International Human Rights Law

[17] 5Rights YouGov poll: Parents' views on internet and child data protection regulation, 2019.
[18] https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services/6-policies-and-community-standards/

22. Failure to extend human rights obligations to online platforms has enabled a culture of impunity, whereby services are excused for knowingly allowing hate speech and harmful content to be targeted at users.[19] For example, a Hungarian platform was deemed to fall outside the scope of the European Convention on Human Rights, despite the fact that it allowed incitements to violence against users to remain on its platform.[20] The first step in creating a multi-level system of liability should therefore be to i**nclude digital service providers within the scope of human rights and children's rights regimes.**

(b) Systemic Duty of Care

23. Service providers should be subject to a 'systemic duty of care' - a forward-looking standard of legal liability that assesses the overall risk of harm presented by a platform's design, user interface and operating mechanisms, rather than imposing liability for any one piece of content.[21] This is in line with the precautionary principle under international law which dictates that, where an online provider designs a service in a way that promotes, algorithmically guides the user towards, primes the publication of, or fails to remove access to, harmful content, then it should bear the distribution of risk where harm arises.

24. The UK's Online Harms proposal adopts a version of such a 'systemic duty of care', to be enforced by Ofcom. But the proposed legislative framework for the Online Safety Bill will not grant any individual an action or remedy in the event that this duty is breached. This standard of liability is not a conventional understanding of a duty of care in negligence law, but is more closely related to a conventional model of statutory regulation. The inability of citizens to bring litigation to enforce their rights to freedom of thought and the freedom to safely access and impart information poses a threat to citizens Article 6(1) ECHR rights to access justice.

(c) Individual Duty of Care

---

[19] As the United Nations Working Group on Online Governance recognised as early as 2004, if today's digital world is the medium of choice for the exercise of democratic citizenship and freedom of speech, then online platforms have assumed the once exclusive role of arbiters of the public sphere. Yet, as the Working Group noted, if online platforms are assuming the role of states, they are not assuming their obligations. Private actors are not bound by international human rights law, guaranteeing citizens the ability to exercise rights only insofar as it is "without interference from a public authority." The Working Group consequently proposed that the human rights regime as reflected in the International Convention on Civil and Political Rights (ICCPR) and the European Convention on Human Rights (ECHR) be extended to apply to online platforms. This was rejected at the time in preference of attempting a softer approach, as reflected in the normative best practice framework established in the 2011 United Nations Guiding Principles on Business and Human Rights ("the Ruggie Principles"). As the UN Special Rapporteur David Kaye noted in 2019, it is now clear that applying such 'soft law' approaches has not changed the behaviour of service providers.

[20] *Magyar Tartalomszolgáltatók Egyesülete and Index.hu Zrt v Hungary Eur* app no 22947/13, ECtHR, 2 February 2013; Ani*mal Defenders International v. UK,* app no. 48876/08, ECHR 22 April 2013.

[21] Following the adoption of the German Network Enforcement Act ("NetzDG") in October 2017, a growing number of EU member states are taking their own initiatives to extend the human rights obligations owed by states to private actors. Although private actors are not subject to human rights obligations, governments are rightly coming to the conclusion that in order to comply with their positive obligations under both the European Convention and the ICCPR, they must take action to regulate online platforms, so that users, including children, can freely access, impart and share information without facing risk of harm. If the European Convention requires governments to give effect to rights that are not "theoretical and illusory but are practical and effective," then it is no longer sufficient that the governments discharge their duty by merely requiring online platforms to abide by notice-and-takedown measures, which put a disproportionate burden of risk on users, including children. Namely, individuals are forced to first experience the infringement of their rights before a remedy can be sought.

25. Services are not passive bystanders to the content that is transmitted on their sites, but rather play an active role in determining the presentation and arrangement of user-generated content, and should be excluded from the scope of immunity provided under Article 19 of the Electronic Commerce Regulations in the UK and Section 230 of the Communications Decency Act 1996 in the US.[22] **Platforms must be liable where they have 'knowledge or control' over information which is transmitted or stored.**

(d) Senior Management and Director liability

26. Individual company directors are the agents who discharge a company's obligation to design, operate and moderate their platform in and manner that does not present undue risk to the end user. Where they fail to do so, the UK should follow the precedent set by s. 198 of the Data Protection Act 2018 and the Financial Services Act 2016, and hold them personally liable.

vi. To what extent should users be allowed anonymity online?

27. Anonymity online can be defined in a number of ways. It can refer to 'pseudonymity' where a person adopts a different persona or hides their identity from the view of other users. It can describe the ability of a person to keep their personal information private from companies, or a state of freedom from government surveillance. These differences are significant because they offer insight into a more nuanced 'layered' approach to anonymity rather than a binary 'on/off'.

28. When considered in relation to freedom of expression, anonymity is generally taken to mean 'pseudonymity' - the ability of a person to remain hidden online by concealing or falsifying their identity or taking on a different identity. Allowing people anonymity online protects their right to privacy and the freedom to express their views without interference.[23] It provides an important safeguard for those living under oppressive or corrupt regimes, for people who are persecuted for their religion or ethnicity, and protects journalists, whistle-blowers and activists when exercising their rights to freedom of expression. For children, anonymity can be a means of exploration, a way of avoiding judgement, stereotypes and assumptions based on age. The ability to be 'invisible' is an important part of play for children. **Where a child can be anonymous online and remain safe, they can enjoy the freedoms that anonymity brings and exercise their rights to privacy and play.**

[22] Historically, tech companies have evaded responsibility for content on their platforms by claiming immunity from civil liability under either Article 19 of the Electronic Commerce Regulations 2003 in the UK or Section 230 of the Communications Decency Act 1996 in the US. At the turn of the millennium, this argument could be justified on the basis that these services functions as 'conduits' and that the services themselves did not exert any degree of knowledge or control over user-generated content. But there is growing consensus among lawmakers both in the US and the UK that these must be revised. This is in light of the fact that the immunity provisions were drafted principally to protect both countries' nascent e-commerce sectors, from fears expressed at the time that they would be held liable for third-party fraudulent payments.

[23] In 2015, the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression stated "encryption and anonymity, and the security concepts behind them, provide the privacy and security necessary for the exercise of the right to freedom of opinion and expression in the digital age." https://www.undocs.org/A/HRC/29/32

29. Although anonymity can provide safeguards for freedom of expression online, it is often abused for nefarious purposes. For this reason, anonymity online often has negative connotations - 'catfishing', fake profiles, grooming, financial scams and so on. When used for the purposes of deception, manipulation or subterfuge, it can cause serious harms to children, including bullying, online sexual abuse, grooming and financial scams. Again, service design also plays a role in encouraging such behaviours, with some digital services designed specifically to allow users to ask and answer questions about each other anonymously. A review of one such app, 'Tellonym', described "questions that are vile, graphically violent, racist, misogynistic and inappropriate… some messages seem to taunt users to consider suicide."

30. There are significant opportunities for companies to employ different layers of anonymity and privacy, and companies have a role to play in ensuring that their users follow their rules. But companies should consider the impact on children of allowing users anonymity on their services or offering 'invisibility' through individual design features. For safeguarding reasons, there may be situations where children must be identified – either by age or by identify – for example to use prohibited services or purchase age-restricted products, but when offered appropriately by services that uphold robust rules and community standards, anonymity can provide important safeguards to freedom of expression

vii. How can technology be used to help protect the freedom of expression?

31. The design of a digital product or service can give rise to harm or help protect against it. When designed in the best interests of children and with regard for their developmental capacity, digital technologies can provide important ways for children to express their views and have their voices heard. Technology should be designed with specific responsibilities for the safety, privacy and rights of the user. To this end, services should reduce algorithmic dissemination of harmful content and re-engineer recommender systems so content that is in breach of rules and/or harmful is not promoted. These are discussed in more detail in our response to question 8.

viii. How do the design and norms of platforms influence the freedom of expression? How can platforms create environments that reduce the propensity for online harms?

32. Many of the risks of the digital world are not at the hands of bad actors but are the cumulative outcomes of common design features and operating processes found across digital services.

<u>How do the design and norms of platforms influence the freedom of expression?</u>

Profiling

33. Tech companies use personal information about their users to analyse and predict behaviour in a practice known as 'profiling'. Profiling is used to make automated decisions and underpins the processes that recommend content and target advertising to users, also known as 'personalisation'. The widespread use of profiling across digital services has a number of detrimental effects on children, but perhaps the most chilling of these - and most overlooked - is its effect on freedom of thought. The algorithms that 'personalise' user experiences make editorial decisions that prioritise certain types of information over others, and rank content not only according to the interests of individual users (inferred from previous engagement or the engagement of others in a child's network) but by criteria shaped by the commercial interests of the service providers and the companies that advertise on their platforms. **The effects of this highly-curated digital ecosystem on a child's freedom of thought have not yet been fully researched, but we can infer from anecdotal evidence and research in other areas of childhood development that it is likely to have a very profound effect on their worldview.**

Automated decision-making

34. Automated decision-making (ADM) systems power features that are ubiquitous across services directed at children, including recommendation loops, nudge techniques and friend/follower suggestions. These are characterised as user 'personalisation', where recommendations are made based on what a user (and 'similar' users) have watched, shared or interacted with previously. ADM recommendation systems are instrumental spreading misinformation online, built to push the most engaging and often most egregious content. These systems can generate a pipeline of increasingly extreme or harmful content,[24] causing users to become locked in an algorithmic echo chamber where misinformation and conspiracy theories are served up in a continuous loop.[25] This is particularly pernicious when combined with other features such as 'auto-play' where video content is not only recommended but begins playing automatically without the user's interaction.

35. Article 17 of the UNCRC states that every child has the right to reliable information from a variety of sources, and that governments must help protect children from materials that could harm them.[26] **Not only do the ADM systems of some services put children at risk of harm, they undermine a child's right to freedom of expression, to access information, their right to participation and their freedom of thought.**

Disappearing content

[24] https://journals.uic.edu/ojs/index.php/fm/article/view/10108/7920
[25] https://www.telegraph.co.uk/technology/2020/10/23/fell-terrifying-conspiracy-theory-wormhole-youtube/
[26] https://www.unicef.org.uk/wp-content/uploads/2019/10/UNCRC_summary-1_1.pdf

36. Features enabling users to create time-bound content that 'disappears' are popular among young people but can create risks from disinhibition and a false sense of ephemerality.[27] 'Disappearing' content also enables the fast spread of misinformation. Such content can be widely shared and viewed before being fact-checked, resulting in false information being spread before being flagged or verified. Young people have expressed concerns about disappearing content and are unsure how to report it once it has 'gone'. However, ephemeral/disappearing messaging should not be misconstrued as a way for children to safely 'make mistakes' or as a solution to indelible posts. Companies will still collect and store data from disappearing content, and screenshots, re-plays, and screen recordings mean there is no guarantee that content actually 'disappears' from the view of other users so that unaccountable information can form part of a child's online identity whilst being unseen. **Disappearing messages have the double effect of encouraging disinhibited, 'consequence-free' behaviour while making such content more difficult to track or truly erase. This has very real impact on a child's right to freedom of expression.**

> "I know from people at my school, that if you have an argument on Facebook, your best friend's got a screenshot, your friend's got a screenshot, and people you don't even know have a screenshot. And it's gone everywhere. And if it's like a video, then loads of people have saved it... it's something that can never really be deleted, because it's happened." - *Young person interviewed by 5Rights*

<u>How can platforms create environments that reduce the propensity for online harms?</u>

37. The attention economy has created an environment in which misinformation flourishes, divisive content is promoted and excessive engagement is the norm for young people. We need to de-toxify the digital ecosystem to create an online space in which young people can access reliable information, where they are not routinely exposed to harmful content or bombarded with targeted advertising and where they can express themselves freely without fear of the thoughts they share being used for commercial exploitation.

38. This detoxification can be achieved in a number of ways. Service providers can:
    - Undertake child impact assessments before rolling out new services, features, or upgrades, to assess how design features may cause risks to accumulate and adapt features accordingly.
    - Introduce friction to prompt users to think before they share with other users.
    - Build algorithms that promote a diversity of views and reliable information, to uphold a child's right to access information.
    - Filter out harmful or inappropriate content (based on official guidance from the Chief Medical Officer, British Board of Film Classification, Department for

[27] 86% of 13-17 year olds use expiring content and use time-bound features to message friends or share amusing content. https://www.childnet.com/ufiles/Youth-perspectives-on-expiring-content---new-youth-research-by-Childnet.pdf

Education and other agreed sources) to young people and ensure that features are age appropriate and evolving with a young person's capacity.
- Set out community rules/standards in a way that demands user compliance as a price of continued access (and at times when users are most likely to engage, at different points in the user journey)
- Uphold community standards and other published terms through robust and consistent enforcement
- Signpost easy to use and robust reporting tools, encourage reporting and offer age appropriate, swift and decisive responses
- Provide accessible and easy routes to challenge and redress in response to content moderation decisions

39. There have been examples of good practice where service providers have adapted their services to address some of risks that threaten freedom of expression online:
    - TikTok has made changes to increase protections for children ahead of the Age Appropriate Design Code (Children's Code) coming into force in September 2021. Accounts belonging to under 16s will now be 'private by default' and only approved followers will be able to comment on videos from these accounts. Users will also be prevented from downloading any videos created by under 16s and direct messaging and live streams will only be available to over-16s.[28]
    - The demotion of clickbait content alongside the promotion of authoritative sources of information, has been a strategy of some online services to tackle the surge in disinformation and misinformation.[29]
    - Trending features have been removed by services in response to scrutiny over curatorial decisions.[30]
    - The creation of a dedicated News Tab where local and national news is displayed has been introduced in a bid to 'highlight original and authoritative reporting'.[31]

    Whilst these examples are indicative rather than complete, they demonstrate that technology can play a significant role in mitigating risks while protecting freedom of expression online.

ix. How could the transparency of algorithms used to censor or promote content, and the training and accountability of their creators, be improved? Should regulators play a role?

40. The digital world is not a singular, neutral space in which encounters are serendipitous, discovery is undirected and experiences are organic. It is a highly-curated environment where deliberate design decisions and automated decision-

[28] https://techcrunch.com/2021/01/13/tiktok-update-will-change-privacy-settings-and-defaults-for-users-under-18/
[29] https://about.fb.com/news/2017/05/news-feed-fyi-new-updates-to-reduce-clickbait-headlines/
[30] https://about.fb.com/news/2018/06/removing-trending/
[31] https://about.fb.com/news/2020/11/launching-facebook-news-in-the-uk/

making systems determine what we see and to a great extent, how we behave. Once described by the founders of Facebook and Twitter as "the digital equivalent of a town square",[32] our newsfeeds and home pages are actually a multitude of unique and individualised town squares. We each have a different engineered reality – and that engineering is largely the work of algorithms. How this affects freedom of thought should not be underestimated. It explains why for most, Trump lost the 2020 election, but for others he won, and why the coronavirus vaccine is for many, the road to herd immunity, and for others, a dangerous ploy to implant microchips into entire populations for the purposes of government surveillance.

41. The impact of a highly-curated digital environment is felt most acutely by young people, with 55% of 12-15-year-olds accessing news via social media sites.[33] They are more susceptible to the damaging effects of misinformation, for example, one in five 16-24- year-olds think there is no hard evidence coronavirus actually exists, compared with fewer than one in twenty 45-75-year-olds.[34] **Whether it is routine exposure to extreme or divisive content or simply being recommended the same types of content over and over again, the algorithms that curate a young person's digital experience have a very profound effect on their freedoms, thoughts, development and wellbeing.**[35]

42. Algorithms also play a role in determining how and by who information is received. 'Algorithmic dissemination' can compound risk by sharing content with larger audiences or changing its meaning and significance through decontextualisation. Content that in itself may not present risk can, through its means of dissemination, become harmful:

> *"Where content is algorithmically disseminated through recommending, this (a) increases its audience, potentially significantly, and (b) typically puts it alongside other, similar content. Rather than speaking of 'harmful' content, then, it is perhaps more accurate to talk about 'potentially problematic' content. That is, content that by itself or when seen only by a relatively small number of people isn't necessarily an issue, but when algorithmically combines with other, similar content or disseminated to a large audience can contribute to systemic problems. Interventions focused on the hosting of content itself miss, to a large extent, issues relating to algorithmic dissemination."* Dr. Jennifer Cobbe, University of Cambridge, and Dr. Jatinder Singh, University of Cambridge[36]

43. If systems are in place that make transparent how and for what purpose data is used and algorithms are built, it will enable regulators, civil society, academics and

[32] Mark Zuckerberg, 'A Privacy-Focused Vision for Social Networking', *Facebook* (6 March 2019) https://www.facebook.com/notes/mark-zuckerberg/a-privacy-focused-vision-for-social-networking/10156700570096634/; Jack Dorsey, *Twitter* (5 September 2018) https://twitter.com/jack/status/1037399093084151808?s=20
[33] https://www.ofcom.org.uk/__data/assets/pdf_file/0013/201316/news-consumption-2020-report.pdf
[34] https://www.kcl.ac.uk/policy-institute/assets/covid-conspiracies-and-confusions.pdf
[35] https://www.technologyreview.com/2020/09/17/1008549/kids-need-protection-from-ai/
[36] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3371830

researchers to identify, understand and mitigate the risks created by design choices and automated decision-making systems. Service providers should combine greater transparency with careful implementation, clear accountability and robust governance processes to reduce the risks presented by automated decision-making systems. **Additionally, it should be an industry norm, supported by regulation to carry out child impact assessments to understand the risks their algorithms pose to young people, then identify appropriate risk-mitigating measures, taking immediate action when algorithms put children at risk.**

44. Transparency is critical if we are to understand the impact of automated decision-making systems on our rights and freedoms. **But transparency alone will not achieve the level of scrutiny required to identify and mitigate risks created by the use of algorithms.** There is also an urgent need for regulatory oversight and enforcement. With transparency must come the required regulatory resources and expertise to understand the effect of ADM systems, to identify appropriate actions and steps to take to mitigate harm, and sufficient enforcement powers to enact those measures. Regulators must have oversight of automated decision-making systems used by services, including the ability to identify and assess the data used to train algorithms (and how that data is collected), to analyse the source code and/or statistical model in use, to assess the impact of the ADM system, and to conduct tests to assess how an algorithm operates in practice and over time.

45. **While transparency and oversight are important, a fundamental cultural shift is needed to stop companies from using automated-decision making systems that power seemingly benign, even useful or 'experience enhancing' features such as auto-play, and content recommendations, until sufficient safeguards are engineered into the design of these systems to protect children's freedoms.** Companies must take responsibility for providing effective training to all staff in the design and governance chain (including developers, engineers, UX designers, product managers, and others) on young people's rights, their vulnerabilities at different stages of development and the range of risks and harms they may experience online as a result. Training should not be restricted to known harms but create a broader understanding of how young people use technology and how technology impacts on their rights and wellbeing.

x. How can content moderation systems be improved? Are users of online platforms sufficiently able to appeal moderation decisions with which they disagree? What role should regulators play?

46. Moderation is a tool of last resort after the community rules of a service have been breached. Moderation should not therefore be considered an affront to freedom of speech, but the Committee should be mindful of the great power that lies in the hands of a few large tech companies, who can choose to silence users, either by removing individual posts or in more extreme cases, 'de-platforming' and shutting down user accounts. Of great concern to 5Rights is the reliance of user-led reporting in the absence of effective automated content moderation. Reporting puts the responsibility for addressing harm on those experiencing it and puts an

unreasonable burden on child users to understand a service's community standards (in which they define what kind of content and behaviour is not allowed), to recognise why something is harmful or inappropriate, and to know how to report it.

47. Content moderation systems can be improved by:
    - Investment in human moderators, relative to the scale and nature of the business.
    - Investment in more sophisticated AI moderation systems to mitigate the spread of harmful content before it reaches new audiences, and to reduce instances of the unlawful take-down of legitimate, informative content.[37]
    - Reducing biases in content moderation so marginalised groups are not censored.[38]
    - Stating clearly in terms of services the purpose of reporting mechanisms, and addressing the abuse of flagging and reporting tools through consistent enforcement.[39]
    - Ensuring published terms (community standards and rules), including how users can report content and appeal a moderation decision, are presented in ways that are age-appropriate and easily understood by a child.
    - Providing effective redress for users to challenge moderation decisions.

Regulators must also be given sufficient enforcement power to hold service providers to account where they fail to provide effective moderation and removal of harmful content.

48. Moderation, reporting and removal of content are important ways of addressing harmful material online. But these mechanisms do little to prevent such material being produced in the first place and focus must be shifted 'upstream' to address the business models and operating systems that allows such content to proliferate. Effective content moderation and removal should not therefore be seen as the extent of a service provider's obligations to address harm that occurs on their services. What must be stressed, is that **the need for effective content moderation, reporting and removal would be greatly reduced if services are designed in ways that do not actively facilitate the spread of harmful content through risky recommendation systems or pernicious design features.**

xi. To what extent would strengthening competition regulation of dominant online platforms help to make them more responsive to their users' views about content and its moderation?

[37] https://www.wired.com/story/coronavirus-social-media-automated-content-moderation/
[38] https://www.theguardian.com/technology/2020/oct/25/instagram-row-over-plus-size-model-forces-change-to-nudity-policy
[39] Reporting mechanisms can be abused if a user with malicious intent reports another user as a form of harassment. There have been cases where a user has created multiple accounts and pages to report posts in a targeted attach against one individual. Disinformation campaigns have also been found to abuse flagging features, with bad actors deliberately flagging posts from figures they disagree with to have them 'cancelled' or their accounts shut down.

49. In a more competitive market, services would compete to offer better alternatives to users who prefer not to share their data, to reduce exposure to distressing material, to respond to user reports more quickly and better uphold community standards. These benefits will be enjoyed by all, but their impact will be felt most by those for who the asymmetry of power between service and user is most pronounced.[40] Children are disproportionately affected by intrusive data collection, aggressive marketing and microtargeting, exposure to harmful content and misinformation. **Strengthening competition regulation of big tech companies will empower young users to make more meaningful choices, give them greater autonomy over their data and create a more diverse digital ecosystem.**

xii. Are there examples of successful public policy on freedom of expression online in other countries from which the UK could learn? What scope is there for further international collaboration?

50. In a global digital environment, international cooperation is crucial for the promotion and protection of the rights of children. As well as ratifying and implementing international instruments to protect children's rights, such as the UNCRC General Comment on children's rights in relation to the digital environment, countries must share knowledge and best practice from national and local level policy interventions. Examples of this are set out below:

51. <u>Move against Section 230</u>
There is growing consensus among lawmakers both in the US and the UK that Section 230 of the US Communication Decency Act, which for years has allowed tech companies to evade responsibility for content on their platforms, must be revised.[41] The scope of section 230 is likely to be reviewed under the new US administration, and significant steps have been made in the UK to protect trade agreements from its influence.[42]

52. <u>Application of international conventions to private companies</u>
In Germany, the Federal Constitutional Court can extend its obligation to uphold international conventions to a private actor if it deems that they operate a form of public space, even if it is commercial in nature. Evaluations of this recently enforced German Network Enforcement Act and a similar law on the sharing of abhorrent violent material in Australia (via the Australian Criminal Code) have demonstrated that when crafted with care, regulations can both address online harms and have a

[40] The concentration of power among a small number of big tech companies has resulted in self-regulation in all but name, with users being deprived of the choice and control to exercise their rights. Young people in particular, for who the digital world is not an option, are faced with a 'take it or leave it' decision when using digital services, where they can either accept terms and conditions set by the one of the powerful service providers or be denied access altogether which given the concentration of so few companies effectively excludes them from the 'public square'.
[41] https://www.justice.gov/ag/department-justice-s-review-section-230-communications-decency-act-1996?utm_medium=email&utm_source=govdelivery
[42] https://www.telegraph.co.uk/news/2021/01/06/duty-care-victory-prevents-trade-deals-watering-laws-protect/

civilising influence on online expression instead of foreshadowing the end of a free and open public discourse.[43]

53. Company Director liability

The UK can learn from New Zealand's approach to holding individual directors accountable where they have failed to ensure their service complies with the regulatory regime. New Zealand's Harmful Digital Communications Act 2015 includes the removal of safe harbour from civil and criminal liability for corporate actors that fail to act on a notice of complaint as to harmful content within 48 hours.

80. Restrictions on targeted advertising

Risks to young people associated with increased exposure to targeted advertising have been addressed through various initiatives across the globe. Indeed, marketing aimed at children has been banned outright in some countries, including Norway, Sweden, Brazil and parts of Canada. Targeted advertising is the central to the business model of many digital services and is in part the reason services providers carry out intrusive data collection practices. If targeting became a less desirable (or illegal) form of advertising for companies, there would be less demand for the data on which microtargeting relies and reduce the rampant processing of children's data. Contrary to the claims of many tech companies whose revenues are generated from advertising, evidence suggests targeting is not always the most effective form of advertising.[44]

We would be very willing to appear in front of the Committee or provide further evidence on any of these points.

For more information, please contact:
Izzy Wick, Policy Lead, 5Rights
izzy@5rightsfoundation.com | 5rightsfoundation.com

[43] see Heidi Tworek and Paddy Leerson, "An analysis of Germany's NetzDG Law," *Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression* (2019), link; Frank Fagan, "Optimal social media content moderation and platform immunities," *European Journal of Law and Economics* (50:2020), link pp. 437-449

[44] In January 2020, the Dutch national broadcaster NPO saw its advertising revenue grow after it stopped targeting advertising across its websites in favour of contextual advertising. (See https://techcrunch.com/2020/07/24/data-from-dutch-public-broadcaster-shows-the-value-of-ditching-creepy-ads/)