Growing up in the online world: a national conversation

Molly Rose Foundation response

May 2026

Summary

Molly Rose Foundation (MRF) welcomes the opportunity to respond to this consultation.

For too long, the needs of children and parents have been poorly served by successive governments, regulators and most of all technology firms. Urgent and decisive action is needed to address the preventable harm faced by young people on social media, gaming platforms, messaging services and AI chatbots.

We strongly support the Government's commitment to take further action on online safety. Parents and young people want to see robust action but also have confidence that the measures adopted will deliver meaningful and long-lasting change.

Why we reject an Australian-style ban

MRF rejects the proposals for an Australian-style social media ban. Multiple studies point to there being significant implementation challenges, with a majority of under 16s retaining access to prohibited platforms.

Research from the University of Chicago also suggests that the high social costs of complying with the ban, coupled with the ease of being able to circumvent social media restrictions, raise significant doubts about whether the threshold needed to achieve and maintain a long-term shift in social norms can be achieved. In fact, that research suggests that compliance with the ban is more likely to fall than rise over time.[1]

In this context, MRF struggles to see a scenario in which a social media ban for under 16s can be successfully introduced in the UK at this time. We believe it would be irresponsible for the government to proceed with a ban unless it can be fully confident that it can be effectively introduced and rigorously enforced.

While there is high superficial support for a ban, polling suggests that UK adults prefer a range of other policy options, not least strengthened regulation.[2] [3] A poorly enforced ban is likely to quickly unravel and introduce additional complexity for parents and young people. Ultimately, if a ban proves ineffective, we risk a majority of children continuing to access platforms where we know harms remain widespread.

Even a well-enforced ban presents a range of potential unintended consequences, not least introducing new barriers to disclosure that may exacerbate the risks that young people face online.

The majority of parents are demanding meaningful change. We should not rush into a measure that risks affording them a false sense of safety, and that delivers a harm-benefit ratio for children that is still to be evaluated and robustly assessed.

[1] Bursztyn, L., Duckworth, A. et al (2026) Why Bans Fail: Tipping Points and Australia's social media man. University of Chicago, Becker Friedman Institute of Economics, Working paper No 2026-57
[2] Molly Rose Foundation (2026) Online safety consultation: the next steps that adults and parents want
[3] Tech at Public First (2026) To Ban or not to Ban

The case for restrictions on high-risk functionalities

We strongly support the Government's proposals to restrict high-risk functionalities and features, including personalised recommender systems, livestreaming and AI chatbots. Recommender systems are the single largest vector through which 13-16 year olds are exposed to harmful content, and in turn the risks of cumulative harm, including from suicide, self-harm and intense depression content.[4]

We therefore support a conditional ban on personalised recommender systems, with platforms having to meet a series of tests if they wish to continue offering personalised recommendations to under 16s. As a minimum, platforms should face clear and unambiguous restrictions on recommending harmful content; new duties to promote a diversity of content types; and a legal requirement to carry content from high quality, trusted sources, such as PSBs.

We also strongly support proposals to restrict a range of addictive design features, including the use of infinite scroll, autoplay features, affirmation functions and alert and push notifications. Much of the justifiable concern currently being expressed by parents stems from the chronic harms being fuelled by the engagement-based design choices adopted by platforms, but which are poorly targeted by the current scope and operation of the Online Safety Act.

MRF also endorses the introduction of a properly enforced minimum user age of access of 13 for social media, messaging services and high-risk gaming platforms like Roblox and Discord. Taken together with the risk-based approach to restricting functionalities set out above, we assert this blended approach of age restrictions and safety-by-design considerations will provide parents and young people with decisive and transformative changes to children's online safety.

For the first time, these measures will also introduce meaningful safety-by-design incentives for regulated firms – if platforms wish to continue to offer products and services to young people, they will have to develop and roll out age-appropriate, safe-by-design versions of their products.

MRF asserts this approach can deliver substantial improvements to children's online safety – and because these measures clearly target the drivers of online harm and attract the broad support of children's groups, civil society and academic expertise, can deliver on the 'confidence premium' that UK adults attach to this next set of Government proposals.

Further action on online safety is required

Though the scope of this consultation is welcome, any new measures must be a downpayment on more ambitious action to decisively protect young people's safety and wellbeing online.

Ultimately, strong and effective regulation remains the most powerful lever to address preventable online harm and tackle the underlying incentives and business models that continue to treat children's safety and wellbeing as little more than a tick box exercise.

The Online Safety Act therefore remains a crucial part of the solution. However, as currently drafted, it is not working as intended.

[4] Molly Rose Foundation (2025) Children's exposure to suicide, self-harm, depression and eating disorder content on social media

In order to truly deliver the reset that children and parents are demanding, the Government must commit to a strengthened and reworked Act.

This attracts the support of the public, with three quarters (73%) of UK adults supporting new legislation to strengthen regulation and better protect children and young people from online harm - stronger than support for an Australian-style ban.[5]

In the current parliamentary session, the Government must commit to a swift and decisive legislative package that can immediately address the most pressing structural barriers to robust and active regulatory enforcement.

Looking ahead, the Government must then make a commitment to a White Paper and announce that a new Act is on the way in the third parliamentary session. This must make provisions for a systemic Duty of Care, reset regulatory incentives in favour of harm reduction, extend the scope of the act to encompass children's wellbeing, and deliver an outcomes- and conduct-based regulatory approach that is commensurate to the size and cash-rich position of the largest companies in the world.

Earlier this year, the Secretary of State promised that 'If we need to be legislating every year on technology, that's what we need to do.' [6] This follows commitments made by Labour to strengthen the Act while in opposition.[7] Government must now keep these promises to deliver the fundamental reset that parents demand and deserve.

[5] Molly Rose Foundation (2026) Online Safety Consultation: the next steps that adults and parents want
[6] Paul, J. (2026) Liz Kendall insists Government will 'legislate every year' to keep up with the pace of technology. Published on LBC, 9th March 2026
[7] Helm, T. (2023) Labour pledges to toughen 'weakened and gutted' online safety bill. Published in the Guardian, 1st January 2023

About this response

This response is structured as follows. Relevant consultation questions are listed in each section's summary.

- **Section 1** explores the risks and benefits of time spent on social media, gaming platforms, messaging services and AI chatbots.
- **Section 2** sets out MRF's overall position on age restrictions, including why we do not support a Australia-style ban, and the case for a blended approach of age restrictions and safety-by-design considerations.
- **Section 3** sets out our position on restricting high risk functionalities. This includes our position on the proactive recommendation of high quality content.
- **Section 4** sets out our position on restricting persuasive design features.
- **Section 5** outlines our priorities for further action to support digital and media literacy.
- **Section 6** sets out our position on other proposals in this consultation, including raising the age of digital consent, time limits, parental controls, and VPN restrictions.
- **Section 7** makes the case for further action to fix and strengthen the Online Safety Act.

Appendix 1 summarises the key themes from MRF's **Open Space** event. As part of the consultation process, this unique delegate-led event brought together Government, Ofcom, civil society, young people and those with lived experience of online harm to discuss how we keep young people safe online. We encourage Government to consider the key themes raised by attendees alongside MRF's response.

Appendix 2 includes supporting information about MRF.

Section 1: The benefits and harms of using social media, gaming platforms, messaging services and AI chatbots

Summary

Social media, gaming platforms, messaging services and AI chatbots all present a range of both harms and benefits to children.

In respect of suicide and self-harm risks, for example, time spent online can have both protective and harmful effects. Though harmful effects currently predominate, this is the result of the current design and operation of online platforms, enabled by ineffective regulation and insufficient safety-by-design incentives on services.

This harm is preventable. In MRF's view, rather than pulling up the drawbridge on children's access, the prize should be eliminating harm so the benefits of technology can be felt by children.

Ambitious action is needed to address the drivers of harm across the stack - including on social media, gaming platforms, messaging services and AI chatbots – rather than a more piecemeal response. This should be targeted not only at the drivers of acute risks to children, but at chronic risks associated with persuasive design and engagement-based business models.

Even if the evidence linking certain aspects of platforms' design and operation to harm is still developing, this should not be a barrier to action, with the actions of services – confirmed by legal disclosures – showing clear intent to maximise time spent and prioritise profit over children's safety and wellbeing.

This section relates to consultation questions 1-3.

The balance of benefits and harms

Evidence shows that time spent online can have both positive and negative impacts on young people, with the overall impact on children at the population level not clear cut. [8] [9]

Though a recent review by the Department of Science, Innovation & Technology found a small correlation between time spent on social media and poor mental health, the study authors

[8] Orben A (2020) Teenagers, screens and social media: A narrative review of reviews and 272 key studies. Social Psychiatry and Psychiatric Epidemiology, 55(4)
[9] Sanders T, et al (2023) An umbrella review of the benefits and risks associated with youth's interactions with electronic screens. Nature Human Behaviour, 8. pp82-99

concluded that they remained unable to determine whether there is a causal impact of time spent online on population-level mental health and wellbeing.[10]

There is, however, strong evidence for individual-level impacts. It is clear that, for some children, time spent online leads to a range of harms, including cyberbullying, sexual exploitation, negative impacts on mental health, and even suicide.

For others, however it can be positive, allowing them to find community, peer support and opportunities for learning and development. Ultimately, while some young people may experience harm from time spent online, others may remain unaffected or benefit from it.

The balance of these impacts is dependent on a range of factors. Crucially, this includes *how* young people use online platforms - including the activities they engage in, the content they consume and the context of their use - which is itself largely determined by the current design and operation of platforms.

It also depends on children's individual strengths and vulnerabilities, with young people responding differently to online risks and hazards.[11] [12]

We have to consider the full complexity of this and also recognise that the balance of harms and benefits are equally valid – we fail children if we only focus on one.

Ultimately, the prize is to address the causes of harm so that the benefits can come to the fore.

Online suicide, self-harm and mental health risks

Almost eight years after Molly's death, evidence shows that suicide, self-harm and mental health risks to children on social media, gaming platforms, messaging services and AI chatbots remain unacceptably high.

Problematic internet use has been identified as a 'common but likely underestimated antecedent' to suicide in young people,[13] with suicide-related internet use reported in 24% of deaths by suicide among young people aged 10 to 19, equivalent to a young life being lost every single week.[14]

Suicide and self-harm related internet use has also been reported in 26% of child hospitalisations relating to self-harm.[15]

There is also a clear relationship between problematic internet use and rates of suicide in groups with protected characteristics. Research shows that suicide-related Internet use is

[10] Department for Science, Innovation & Technology (2026) Understanding the impact of smartphones and social media on young people
[11] Ibid.
[12] Office of the Surgeon General (2023) Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory
[13] Susi, K. et al (2023) Research Review: Viewing self-harm images on the Internet and social media platforms: systematic review of the impact and associated psychological mechanisms. Journal of Child Psychology and Psychiatry, 64(8). pp1115–1139
[14] Rodway, C et al (2022) Online harms? Suicide related online experience: a UK wide case series study of young people who die by suicide. Psychological Medicine, 53 (10), pp1–12
[15] Padmanathan, P (2018) Suicide and Self-Harm Related Internet Use: a Cross-Sectional Study and Clinician Focus Groups. Crisis, 39(6), pp469-478

recorded more frequently in the death by suicide of girls, and in cases affecting adolescents identified as LGBTQ+.[16]

In this subsection we discuss three key drivers of online suicide and self-harm risks – the adverse effects of harmful content, online suicide and self-harm offences, and the role of AI chatbots.

The adverse effects of online suicide, self-harm and depression content, and the design choices that promote it

There is growing evidence of the relationship between exposure to harmful content online and resulting suicide and self-harm risks.

While engaging with suicide, self-harm and intense depression material has been found to have both harmful and protective effects, there is evidence that the harmful effect currently predominate.[17] The impacts of engaging with this material may include:

- Increases in the frequency and/or severity of suicide and self-harm ideation and behaviours. Arendt et al (2019) found that one third of participants in their study carried out the same or similar types of self-harm after observing it on Instagram.[18]
- Engagement behaviours such as sharing, liking or commenting on suicide and self-harm content may reinforce the creation and sharing of self-harm images, and in turn encourage further harmful behaviours;[19] while engaging with material may result in a range of emotional, cognitive and physiological impacts, for example increased rumination, which may go on to trigger or exacerbate themes of despair, hopelessness and in some cases suicide and self-harm ideation.
- Personalised recommender feeds and other engagement-based design features may result in an 'assortative relating' effect, with young people experiencing suicide ideation or thoughts of self-harm being more likely to identify and build relationships with other users experiencing similar actions and thoughts.[20]
- Online communities of practice may form, which in some cases may offer protective effects but that can also result in young people experiencing an exaggerated perception of suicide, self-harm and intense depression behaviours; a risk that self-harm is seen as normalised or beneficial coping mechanism; and a risk that these online cultures may inadvertently preclude offline or expert oriented forms of help seeking (cementing a sense that those who do not self-harm ''cannot understand)'.[21]

[16] Ibid.
[17] Susi, K et al (2023) Research Review: Viewing self-harm images on the Internet and social media platforms: systematic review of the impact and associated psychological mechanisms. Journal of Child Psychology and Psychiatry, 64(8). pp1115–1139
[18] Arendt, F et al (2019) Effects of exposure to self-harm on social media: evidence from a two-way panel study among young adults. New Media and Society, 21. pp2422–2442
[19] Ibid.
[20] Ibid.
[21] See, for example, Lavis, A et al (2020) #Online harms or benefits? The graphic analysis of the positives and negatives of peer support around self-harm on social media. Journal of Child Psychology and Psychiatry, 61. pp842–854

The availability and reach of harmful content

MRF research conducted immediately before the Online Safety Act's Protection of Children Codes came into effect in July 2025 identified that harmful content remains readily accessible and discoverable on major social media platforms.

Using a well-established safety testing methodology, our analysis identified that harmful suicide, self-harm and depression content was present in overwhelming quantities on major social media platforms, with recommender feeds and other engagement-based features actively pushing this to teenage accounts.[22]

After a short period of engagement with popular suicide, self-harm and depression material via an 'avatar' account set up in the guise of a 15-year-old girl, 96% of algorithmically recommended videos on TikTok's For You Page and 97% of short-form videos on Instagram Reels were found to be potentially harmful. On Reels, this include 8% of recommended content promoting or glorifying suicide or self-harm, and 20% referencing suicide methods.

This content was reaching vast audiences. Harmful posts algorithmically recommended on TikTok's For You Page, for example, had been liked by an average of over 303,000 accounts, with one in ten (9%) liked at least a million times. On Instagram, harmful reels attracted an average of 226,000 likes.[23]

Children's exposure to harmful content on major social media platforms

Building on the above research, in Spring 2025 MRF conducted a large-scale survey to understand teens' exposure to potentially harmful content on six major social media platforms across four themes: suicide, self-harm, depression and eating disorders. 1,897 children aged 13 17 from across the UK took part, covering a range of backgrounds and characteristics.[24]

Our analysis identified that – in the previous week – over a third (37%) of teens had seen at least one kind of suicide, self-harm, depression or eating disorder material likely to be classed as either Primary Priority Content or Non-designated content under the Online Safety Act. This included 5% of 13-17 year olds who had seen *content that encourages or promotes suicide,* 6% who had seen *content that encourages or promotes self-harm,* and 12% who had seen *content that encourages or promotes eating disorders*.

Certain groups of children known to be more vulnerable to online harm were exposed at even higher rates. Around half (49%) of girls reported seeing at least one type of content likely to be classed as PPC or NDC in the last week, compared with a quarter (25%) of boys. Similarly, two thirds (68%) of children with low wellbeing reporting seeing PPC or NDC, compared to one in five (20%) with high wellbeing. Two in five (43%) children with SEND reported that they had seen these content types, compared to 35% of those without.

Patterns of exposure to harmful content also suggested that many children were likely to be at significant risk of experiencing cumulative harm. Many children were encountering content likely to be classed as PPC or NDC repeatedly. 27% of those who had seen *content that encourages or promotes suicide*, for example, had seen this more than 10 times on a platform in

[22] Molly Rose Foundation (2025) Pervasive by Design
[23] Ibid.
[24] Molly Rose Foundation (2025) Children's exposure to suicide, self-harm, depression and eating disorder content on social media

the last week. Many were also seeing high risk content in combination with other forms of related material likely to increase their risk of experiencing cumulative harm.

Recommender feeds remain the single biggest vector for exposing young people to harmful content

MRF research has repeatedly confirmed that, over eight-years since 14-year-old Molly was algorithmically bombarded with over 2000 items of harmful content on Instagram in the months leading up to her death, recommender feeds remain the single biggest vector for exposing young people to harmful content.

Our survey of 13-17 year olds identified that over half of children (between 53% and 57% depending on the theme of harmful content) who reported encountering suicide, self-harm, depression or eating disorder content had done so on recommender feeds such as Instagram's Explore and Reels pages, or TikTok's For You Page. This was considerably higher than any other product surface.[25]

MRF's concern around the impact of recommender feeds is shared by the UK public, with nine in ten (91%) of UK adults saying that they are concerned about suicide and self-harm content being recommended to children and young people on social media.[26]

Suicide and self-harm offences that span social media, gaming platforms and messaging services

MRF is profoundly concerned by the growing threat that online suicide and self-harm offences pose to young people. This is a concern we share with parents, with MRF polling finding that an overwhelming majority (93%) are concerned about children being groomed into acts of suicide and self-harm.[27]

We are particular concerned by the profound risks to children posed by Com Networks – fluid online networks where offenders collaborate and compete to cause harm across a broad spectrum of criminality, including coercing vulnerable minors into suicide and self-harm.

The threat posed by these groups has been the subject of at least five advisories by global law enforcement agencies. In March 2025, the National Crime Agency reported that reports of this threat had increased six-fold in the UK from 2022-2024.[28] Most recently, Europol identified these groups as 'an extremely serious threat to children and society as a whole.' [29]

A threat assessment conducted by Resolver Trust and Safety in partnership with MRF highlighted the complexity of this threat, with the harms facing children increasingly 'hybrid, interconnected and global'. Rather than representing a siloed threat profile, for example, perpetrators are known to commit 'the widest and most extreme range of illegal and harmful behaviours', while also preying on a range of intersecting vulnerabilities. The Com ecosystem is

[25] Molly Rose Foundation (2025) Children's exposure to suicide, self-harm, depression and eating disorder content on social media
[26] Molly Rose Foundation (2026) Online Safety Consultation: the next steps that adults and parents want
[27] Ibid.
[28] National Crime Agency (2025) Sadistic online harm groups putting people at unprecedented risk, warns the NCA
[29] Europol (2026) Internet Organised Crime Threat Assessment (IOCTA) 2026

also decentralised, with new subcultures and clusters quickly emerging and then fragmenting, necessitating that regulation keep pace with a constantly evolving set of risk signals.[30]

Crucially, the Threat Assessment identified that these groups operate and target victims across a wide range of online platforms, with risks highest on gaming platforms, messaging apps and livestreaming services. Ultimately, 'any platform enabling user-to-user connectivity is at risk of exploitation'. [31]

Perpetrator activity also spans across multiple platforms and 'exploits gaps between systems'. A common pathway to harm, for example, might be a perpetrator making initial contact with a victim via a forum, moving them to a direct messaging platform, and then to a livestream to capture acts of violence and abuse. Perpetrators are also known to adapt their behaviour and move platforms if they are unable to operate due to new restrictions.[32]

As such, any measure to protect children that is focused solely on social media or a small subset of platforms is poorly calibrated to the threat profile posed by these groups.

AI chatbots

There is growing evidence that poorly designed AI chatbots pose a substantial risk to the mental health and wellbeing of young people.

Evidence shows that AI chatbots are now extensively used by children.[33] [34] They are also more likely to be used by vulnerable children, with 26% of vulnerable children saying they would prefer to talk to a chatbot than a real person.[35]

They are also increasingly being used for mental health support by both children and adults, including by those experiencing suicide or self-harm ideation.[36] [37] 41% of UK teens feel like people their age are relying heavily on AI for emotional support or help with emotional issues, while one in seven (14%) say they use AI to discuss things they don't feel they can talk to anyone else about.[38] Despite this, many chatbots display a lack of any basic safeguarding measures.

In particular, chatbots consistently feature sycophantic and affirming prompts that can reinforce rather than challenge suicidality and self-harm ideation.[39] Where teens have confided

[30] Resolver Trust and Safety, in partnership with Molly Rose Foundation (2026) Weaponised loneliness: Critical Harm Intelligence Briefing
[31] Ibid.
[32] Ibid.
[33] UK Safer Internet Centre (2026) Smart tech, safe choices: Exploring the safe and responsible use of AI
[34] Internet Matters (2025) Me, myself and AI: Understanding and safeguarding children's use of AI chatbots
[35] Ibid.
[36] McBain, R (2025) Teens are using chatbots as Therapists. That's alarming. Published in New York Times, 25th August 2025
[37] Mental Health UK (2025) Over one in three using AI Chatbots for mental health support, as charity calls for urgent safeguards
[38] UK Safer Internet Centre (2026) Smart tech, safe choices: Exploring the safe and responsible use of AI
[39] DeFreitas, J et al (2025) Emotional Manipulation by AI Companions. Harvard Business School Working Paper No. 26-005 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5390377

in chatbots of mental health crises, there have been examples of their feelings and impulses being affirmed without appropriate safeguarding mechanisms in place – leading to tragedy.[40]

Research from University of Cambridge suggests that children may be particularly vulnerable to inappropriate or harmful prompts when using AI chatbots due to 'empathy gap' in chatbot responses to abstract, emotional, and unpredictable conversations.[41] Children are also more likely to perceive chatbots as quasi-human and trustworthy.[42]

Research also shows that existing safeguards can quickly be circumvented to allow chatbots to produce instructional self-harm and suicide related content.[43] The Centre for Countering Digital Hate, for example, identified that ChatGPT 5 could be made to provide users with methods for self-harm, particularly following multi-turn conversations. [44]

In this context, MRF polling found that 85% of adults are concerned about children being given harmful or inappropriate mental health advice when using AI chatbots (with over half of adults very concerned).[45]

These risks are compounded by chatbots' wider use of manipulative and engagement-based responses that incentivise children to spend excess time and place undue trust in products that are currently often unsafe for use. For example, experiments with 3,300 nationally representative US adults showed that 'manipulative farewells' ('would you like me to do x'? why don't we stay and talk about y'?) can boost post-goodbye engagement by up to fourteen times.[46] These are deliberate design choices, exemplified by OpenAI instructing its designers in November 2025 to increase daily active users by 5% by the end of the year.[47]

Use of AI Chatbots may also drive more chronic risks, with evidence of children spending extended periods of time on AI companion apps,[48] and some evidence that teens' use of AI may reflect features of behavioural addiction, including disrupted sleep, academic struggles and strained relationships.[49]

[40] Reiley, L (2025) What Chat GPT Told My Daughter before She Took Her Life. Published in New York Times, August 24th 2025
[41] Kurian, N. (2025) 'No, Alexa, no!': designing child-safe AI and protecting children from the risks of the 'empathy gap' in large language models. *Learning, Media and Technology*, *50*(4), pp621–634.
[42] Ibid.
[43] Schoene and Canca (2025) For argument's sake, show me how to harm myself!: Jailbreaking LLMS in Suicide and Self-harm contexts
[44] Centre for Countering Digital Hate (2025) The illusion of AI safety: Testing OpenAI's new Safe Completions approach to chatbot safety
[45] Molly Rose Foundation (2026) Online Safety Consultation: the next steps that adults and parents want
[46] DeFreitas, J et al (2025) Emotional Manipulation by AI Companions. Harvard Business School Working Paper No. 26-005
[47] Hill, K et al (2025) What OpenAI did when ChatGPT users lost touch with reality. Published in New York Times, November 23rd 2025
[48] DeFreitas, J et al (2025) Emotional Manipulation by AI Companions. Harvard Business School Working Paper No. 26-005
[49] Namvarpour, M et al (2025) Understanding Teen Overreliance on AI Companion Chatbots Through Self-Reported Reddit Narratives. Published in arXiv e-print

Chronic harms arising from persuasive design and engagement-based business models

Much of the justifiable concern currently being expressed by parents stems from the chronic harms associated with the addictive and persuasive design features that underpin engagement-based business models. This includes the opportunity costs of excess screen time, compulsive behaviour, damaged relationships, and other negative consequences for health and wellbeing.[50]

As it stands, these harms are poorly targeted by the current scope and operation of the Online Safety Act, which focuses on acute risks, such as illegal content and activity, and content that is harmful to children. This reflects the legislative context in which the Act was developed – and it was right that Parliament afforded priority to establishing baseline protections against the most pressing and immediate of threats.

Though evidence for chronic harms associated with persuasive design is still developing, research shows a range of potential negative effects.

Both children and parents share concerns about the impacts of excess screentime. Recent Ofcom research finds that almost four in ten (37%) 8-17 year-olds think they spend too much time on screens, rising to over half (56%) of those with mental health conditions. This concern is shared by parents, with over half (55%) of parents of 8-17 year-olds feeling their children spend too much on screens.[51] A 2019 study also identified that one in four young people experience 'problematic smartphone use' – or difficulties controlling or regulating smartphone use – which can be linked to a range of negative wellbeing impacts.[52]

Some children also report experiencing opportunity costs due to time spent on online platforms. According to recent Internet Matters research, 46% of children report continuing to play the same games or watch the same TV shows or films even while not enjoying them, while over a third (37%) of parents reported that their child turns down opportunities to meet with friends so they can stay in on their phone, computer or games console.[53]

Persuasive design also has clear implications for children's agency over their online experiences. Agency is widely recognised as fundamental psychological need, and yet digital design often serves to frustrate the control children have over their online experiences.[54] When children do not feel able to regulate their online behaviour, this can have negative impacts on their wellbeing.[55]

Other chronic impacts that may be linked to time spent online include disrupted sleep and attention problems.[56] Though social media and other platforms can have positive impacts on children's development through enhanced education and learning, excessive screentime can

[50] 5Rights Foundation (2023) Disrupted Childhood: The cost of persuasive design
[51] Ofcom (2026) Children and Parents: Media Use and Attitudes Report
[52] Ndayambaje, E & Okereke, PU (2025) The Psychopathology of Problematic Smartphone Use (PSU): A Narrative Review of Burden, Mediating Factors, and Prevention. Health Sci Rep, 8(5)
[53] Internet Matters (2026) Children's Wellbeing in a Digital World: Year Five Index Report 2026
[54] Skeggs, A & Orben, A (2025) Social media interventions to improve well0being. Nature human behaviour, 9, pp1079-1089
[55] Internet Matters (2022) Intentional use: How agency supports young people's wellbeing in a digital world
[56] U.S. Surgeon General (2023) Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory

also be linked to negative impacts on cognitive development, academic outcomes and language development.[57]

How persuasive design and chronic harms are 'baked in' to engagement-based business models

The chronic impacts of time spent on social media, gaming platforms, messaging services and AI chatbots are the direct result of business models that prioritise maximising children's engagement over their safety and wellbeing. By employing persuasive design features that are designed to be addictive, services aim to increase the time spent or level of active engagement with a service in order to increase revenue generated through advertising.

These deliberate design choices draw on behavioural science to influence children and encompass a wide range of features not referenced in this consultation, including any features or user interfaces that capture and hold users' attention, create habitual behaviours, or 'nudge' them to engage with content or interact with others in a certain way.[58] Some of these features are clear examples of 'dark patterns' - intentional design choices that manipulate or deceive users into making decisions that are not in their best interest and compromise user safety, privacy or wellbeing.[59]

Evidence shows that children may be particularly vulnerable to being influenced by these features. This is in part due to children's developing brains, including the fact that they are less capable of resisting impulses and regulating their behaviour.[60]

Though the evidence linking persuasive design features to chronic harms is still emerging, legal disclosures demonstrate that platforms are aware of the potential negative impacts of these features on children, revealing clear intent to maximise engagement and profit over children's safety and wellbeing.

Recent landmark trials against Meta and YouTube, for example, found that features like infinite scroll were 'designed to be addictive', and contributed to compulsive use and serious mental health harms to users.[61] In parallel, a separate New Mexico case concluded that Meta's platform design practices harmed children's wellbeing and concealed known risks.

Platform's prioritisation of engagement over wellbeing is best illustrated by Meta's 'Project Daisy'. In 2019, Facebook and Instagram re-designed their affirmation features to hide 'like counts' on users' posts. Although internal evidence showed that this change improved teens' wellbeing by reducing social comparison and anxiety around likes, Meta chose not to make the feature a default because it reduced engagement metrics and was projected to lower revenue by an average of 1%.[62]

Engagement-based design and acute harms

[57] Muppalla, S et al (2023) Effects of Excessive Screen Time on Child Development: An Updated Review and Strategies for Management. Cureus, 15(6)
[58] Ibid.
[59] Behavioural Insights Team (2025) Behavioural Audit of Online Services: Key Findings report
[60] Hartley, CA & Somerville, LH (2015) The neuroscience of adolescent decision-making. Current Opinion in Behavioral Sciences, 5, 108–115. https://doi. org/10.1016/j.cobeha.2
[61] Hays, K (2026) Campaigners welcome Meta and YouTube's defeat in landmark social media addiction trial. Published in BBC News, March 2026
[62] Mirza, R (2023) Case Study on Online Youth Harms – Project Daisy

As well as driving chronic harms, addictive and persuasive design features can also drive children's exposure to more acute risks.

As explored below, this includes how autoplay and infinite scroll features encourage children to spend excess time on the algorithmic feeds that are responsible for the majority of children's exposure to harmful content, and how personalised notifications may direct children to engage with harmful content.

MRF research continues to identify a range of other engagement-based design features that increase the overall risk profile associated with suicide, self-harm and depression risks. This includes new AI -generated search prompts that appear when a user watches or engages with recommended content on TikTok's For You page. Our analysis found this feature enables users to immediately and more readily than ever enter 'rabbit holes' of harmful content, including extreme suicide and self-harm material, alongside material that could exacerbate the risk of cumulative harm. Other examples include 'red dots' used to signal the popularity of potentially harmful content, and ephemeral stories and broadcast channels that use scarcity principles to drive engagement.[63]

Key persuasive design features linked to both acute and chronic harms

As explored in more detail in Section 4, MRF encourage Government to take action against a broad range of persuasive design features and dark patterns. In this section, we set out our concerns about three key features.

Autoplay and infinite scroll

Both autoplay and infinite scroll are deeply embedded within engagement-based business models. Often used in combination with recommender systems, these features are designed to create a 'seamless' user experience, remove any cues to disengage, and ultimately to maximise engagement. One study describes how the design of TikTok's For You Page encourages high levels of engagement and commitment as part of a 'flow experience', serving up an 'endless, hard-to-anticipate, flow of auto-looped videos.[64]

These features play a key role in encouraging compulsive and excess use of social media, exploiting children's more limited ability to regulate their behaviour and making it difficult for them to disengage from platforms.[65] [66]

Given that they incentivise high levels of use on recommender feeds, these features are also closely linked to exposure to harmful content, with MRF research identifying that over half of 13-17 year olds who encountered suicide, self-harm, depression and eating disorder content had done so on recommender feeds, such as TikTok's For You Page.[67]

Affirmation features

[63] Molly Rose Foundation (2025) Pervasive by Design
[64] Conte G et al (2025) Scrolling through adolescence: a systematic review of the impact of TikTok on adolescent mental health. European Child & Adolescent Psychiatry, 34(5)
[65] American Psychological Association (2026) Potential Risks of Content, Features, and Functions: A closer look at the science behind how social media affects youth.
[66] Behavioural Insights Team (2025) Behavioural Audit of Online Services: Key findings report
[67] Molly Rose Foundation (2026) Children's exposure to suicide, self-harm, depression and eating disorder content on social media

MRF has strong concerns about the chronic and acute risks that affirmation features pose to children and young people. Affirmation features are a key aspect of engagement-based business models, with 'likes', 'scores' and other forms of quantification activating reward learning systems by signalling positive social feedback, thereby encouraging increased levels of posting and other forms of engagement.[68] Evidence shows that children are particularly susceptible to these features, with adolescents brains wired to be more aware of social comparison, and seek out attention and affirmation from their peers.[69]

Recent research also finds that affirmation features can be linked to negative mental health impacts, with increased sensitivity to social media rewards such as a 'likes' predicting declines in mental health over time, with users who update their posting behaviour in response to likes reporting worse mental health over time.[70] As above, Meta's 'Project Daisy' found that removing these features could to improvements in children's wellbeing.[71]

MRF also has specific concerns about the risk posed by 'streaks', with many children feeling extreme social pressure to 'keep' up with a Snapchat streak, with knock-on consequences for compulsive use and wider wellbeing.[72] [73]

Affirmation features can also contribute to more acute risks. In the context of suicide and self-harm content, for example, likes and reposts may provide validation from other users and reinforce negative thought patterns or behaviours.[74] MRF research, for example, found that harmful suicide, self-harm and depression posts algorithmically recommended on TikTok's For You Page had been liked an average of over 303,000 times, with harmful Instagram reels attracted an average of 226,000 likes.[75]

Notifications and other prompts

Notifications, alerts and other prompts are a key driver of children's engagement with online platforms, encouraging them to return to a service on a regular basis.

Though some notifications offer benefits, as currently designed, high frequency and intrusive notifications can have a clear negative impact on children's wellbeing, with evidence that they can disrupt children's focus and undermine academic performance and productivity. They can also contribute to compulsive excessive use, exploiting the brain's reward system and fostering habitual usage.[76]

---

[68] Davidson, BC, Turner, G, Gunshera, LJ & Orben, A (2026) Social Media Reward Sensitivity and Depressive Symptoms Over Time. The Longitudinal Relationship between Social Media Reinforcement Learning and Mental Health

[69] American Psychological Association (2026) Potential Risks of Content, Features, and Functions: A closer look at the science behind how social media affects youth.

[70] Ibid.

[71] Mirza, R (2023) Case Study on Online Youth Harms – Project Daisy

[72] 5Rights Foundation (2023) Disrupted Childhood: The cost of persuasive design

[73] Essen, C & Ouytsel, J (2023) Snapchat streaks – How are these forms of gamified interactions associated with problematic smartphone use and fear of missing out among early adolescents? Telematics and Informatics Reports, 11

[74] Susi, K et al (2023) Research Review: Viewing self-harm images on the Internet and social media platforms: systematic review of the impact and associated psychological mechanisms. Journal of Child Psychology and Psychiatry, 64(8). Pp1115–1139

[75] Molly Rose Foundation (2025) Pervasive by Design

[76] Liu, T et al (2024) The Impact of Social Media on Children's Mental Health: A Systematic Scoping Review

They can also drive exposure to more acute risks, with MRF research identifying that – eight years after Molly received personalised recommendation emails that encouraged her to view harmful suicide, self-harm and depression material – platforms continue to use dark patterns and scarcity principles to prompt teen users to log onto to platforms and view potentially harmful content.[77]

The benefits of time spent on social media, gaming platforms, messaging services and AI chatbots

Though many children experience harm, evidence also shows that time spent on social media, gaming platforms, messaging services and AI chatbots can provide a range of benefits for young people. This includes providing positive community; connection with others who share identities, abilities and interests; protective effects for mental health; and opportunities for learning and development.

Recent Internet Matters research, for example, found that children report benefits of time spent online across the developmental, emotional, physical and social domains of wellbeing, including using the digital world to stay in contact with friends and family (83%), to discover new hobbies and interests (74%) and find out what they would like to do in the future (75%). Overall, though children reported a mix of impacts on their wellbeing, children reported the positives of online life still outweigh the negatives.[78]

In this context, it is crucial that we do not dismiss the clear benefits that online life offers to many young people. Blunt solutions like an Australian-style ban risk a range of serious unintended consequences, particularly for children from marginalised or vulnerable backgrounds.

Social connection and finding community

One of the key benefits of time spent online is how it supports positive social connection, with evidence that social media, gaming and messaging platforms can help adolescents maintain existing links with peers, build new connections, and foster a sense of belonging, particularly during 'active' engagement with social media.[79]

Research by the Mental Health Foundation found that three quarters (76%) of young people reported feeling very or somewhat connected with others through online communities, while two thirds (63%) had been in an online community which had made them feel more confident or supported in who they are.[80]

Benefits for children with one or more protected characteristics

The benefits of time spent online may be particularly apparent for marginalised groups of young people, with time spent online offering benefits around identity formation, peer support and

[77] Molly Rose Foundation (2025) Pervasive by Design
[78] Internet Matters (2026) Children's Wellbeing in a Digital World: Year Five Index Report 2026
[79] Nagata JM et al (2024) Health Benefits of Social Media Use in Adolescents and Young Adults. Current Pediatrics Research, 13(1)
[80] Mental Health Foundation (2025) Online communities, safety and young people

acceptance. This is, however, a complex picture, with marginalised groups also more likely to be exposed to online risks.

For example, a systematic review identified that social media use may support the mental health and wellbeing of LGBTQ youths through enabling peer connection, identity management, and social support.[81]

There is also evidence of positive impacts on children in minority ethnic groups, with research by Common Sense Media finding that seven in ten (71%) adolescent girls of colour who use TikTok or Instagram had encountered positive or identity-affirming content related to race at least monthly.[82]

Research also indicates that social media use can offer wellbeing benefits to children with Special Educational Needs and Disabilities, making it easier for young people to find a sense of belonging, connect with peers who share their interests, self-regulate, and learn.[83] [84] Looked after children can also experience benefits, with social media and other platforms supporting young people living in care to maintain healthy relationships with their birth families, make new connections, and ease transitions between placements.[85]

Development and participation

Time spent on social media, gaming platforms and other services can offer powerful opportunities for development – to learn, play, be creative and access information and support that may not be available in children's offline lives.[86] This includes opportunities for the development of digital and media literacy skills that are vital for children to thrive in a digital economy.

Time spent on these platforms can also support a wider range of children's rights, including their rights to participation - allowing them to form opinions, express themselves, and freely join social and political activities.[87] In recent polling, four-fifths (40%) of 10-16 year-olds say that social media helps them to advocate or speak out on issues important to them.[88]

Protective effects for mental health, suicide and self-harm

Research suggests that, while online platforms as currently designed and operated pose unacceptable risks to young people's mental health, they can also support positive outcomes.

For example, a major review of the impact of viewing self-harm content online found that, though harmful effects predominate, this can also have protective effects, including self-harm

[81] Berger, M. et al. (2022) Social Media Use and Health and Well-being of Lesbian, Gay, Bisexual, Transgender, and Queer Youth: Systematic Review. *Journal of Medical Internet Research*
[82] Common Sense Media (2023) Teens and Mental Health: How Girls Really Feel About Social Media
[83] SEND Network (2024) Screen time: how it impacts the wellbeing and learning of children with SEND
[84] Child Mind Institute (no date) Neurodivergent Kids and Screen Time
[85] University of East Anglia (2020) The benefits of social media for young people in care
[86] OECD (2025) From playgrounds to platforms – childhood in the digital age
[87] Digital Futures Commission and 5Rights Foundation (2023) Child Rights by Design
[88] Girlguiding (2026) Social media companies need to do more to protect young people

mitigation or reduction, promotion of self-harm recovery, and encouraging social connection and help-giving.[89]

Similarly, while online mental health communities can have negative impacts on young people's wellbeing, they can also offer a sense of validation and belonging, helping young people to regain control over their wellbeing.[90]

Time spent online can also help young people access support, with evidence that online mental health interventions may help promote help-seeking behaviours, mental health destigmatisation and health education.[91] [92]

Our ambition must be to address the causes of harm so that these benefits can come to the fore.

[89] Susi, K. et al (2023) Research Review: Viewing self-harm images on the Internet and social media platforms: systematic review of the impact and associated psychological mechanisms. Journal of Child Psychology and Psychiatry, 64(8). pp1115–1139

[90] Stoilova, M et al (2021) Adolescents' mental health vulnerabilities and the experience and impact of digital technologies: A multimethod pilot study. London School of Economics and Political Science and King's College London.

[91] Office of the Surgeon General (2023) Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory

[92] Nagata JM et al (2024) Health Benefits of Social Media Use in Adolescents and Young Adults. Current Pediatrics Research, 13(1)

Section 2: Age restrictions on access to social media, gaming platforms, messaging services and AI chatbots

Summary

Age-based restrictions are an important aspect of ensuring children's safety and wellbeing online.

However, MRF does not support an Australia-style social media ban for under 16s. Though we recognise the appealing simplicity of ban, it may ultimately cause more harm than good.

Multiple studies now point to significant implementation challenges in Australia, with a majority of under-16s retaining access to prohibited platforms, and compliance likely to fall over time. MRF struggles to see a scenario in which a ban could be implemented or enforced meaningfully differently in a UK context, and believe it would be irresponsible for Government to proceed with a ban that may provide parents with a false sense of safety in the short term, and then ultimately unravel.

Even a well-enforced ban presents a range of unintended consequences, and we should not rush into a measure whose harm-benefit ratio has yet to robustly evaluated.

Instead, MRF supports a blended approach of age restrictions and safety-by-design considerations, namely:

- the introduction of a properly enforced minimum user age of access of 13 for social media, messaging services and high-risk gaming platforms like Roblox and Discord
- risk-based age ratings based on a platform's functionalities
- by implication, the robust option of a set of safety by design incentives

Rather than being narrowly targeted at a subset of platforms, this risk-based approach can apply across the stack – encompassing social media, gaming platforms and messaging services. Our proposals relating to a conditional ban on AI chatbots for under-16s are set out in Section 3.

This approach also aligns with parents and children's priorities. Although they are profoundly concerned about children's online safety, parents consistently prefer measures that enable children to continue using social media safely over blunt access restrictions.[93] Similarly, two thirds (69%) of 10-16 year olds say they would prefer to know if a platform is safe than be banned from it.[94]

[93] Tech at Public First (2026) To Ban or not to Ban
[94] Girlguiding (2026) Social media companies need to do more to protect young people

Parents also want confidence that solutions will work – a need clearly met by an approach that attracts expert support and takes an evidence-led approach to addressing known drivers of harm.

This section relates to consultation questions 4-7, 21-25, and 31-35

Restricting services by age

MRF strongly supports measures to introduce a minimum age for accessing social media, messaging services, and high-risk gaming platforms like Roblox and Discord.

As it stands, the Online Safety Act necessitates that user-to-user services apply their terms of service consistently. In most cases, this means upholding a minimum user age of 13.

However, it is clear that Act has largely failed to deliver adequate enforcement of these measures, building on the demonstrable failure of the ICO to openly enforce the provisions of its own Children's Code.

The absence of a mandatory requirement to use highly effective age assurance represents a striking omission in the Act. As a result, evidence suggests that the vast majority of under 13s continue to use one or more social media accounts.[95]

A properly enforced minimum age of 13 would bring the legal and regulatory requirements on relevant user-to-user services in line with public expectations. As set out below, this should be accompanied by new risk-based age ratings based on a platform's functionalities.

However, MRF strongly opposes proposals to raise the minimum user age from 13 to 16.

As set out below, an Australian-style ban introduces significant disincentives for companies to comply with restrictions, meaning it is likely to give parents a false sense of safety while continuing to put children at risk. It also comes with a range of unintended consequences while failing to deliver on the 'confidence premium' that parents attach to upcoming Government action, prioritising access restrictions over an appropriate mix of access and safety-by-design considerations.

Why an Australia-style under-16s social media ban is not the right approach

Early evidence from Australia raises significant questions around the implementation of an outright ban, with clear implications for whether this approach can offer parents in the UK the decisive and effective action they demand and deserve. It also risks wide range of potential unintended consequences.

Early evidence from Australia

[95] Ofcom (2025) Children and parents: media use and attitudes report 2025

In March 2026, Molly Rose Foundation undertook the first large scale polling of Australians aged 12-15 on the country's social media ban.[96] Our results suggest there are significant questions about the initial effectiveness of Australia's approach.

Multiple other studies, including the Office of the e-Safety Commissioner's own data, similarly suggest that - at least initially - Australia's ban has not led to a majority of young people losing access to their social media accounts.

Our findings suggest that:

- A clear majority of Australian 12 to 15-year-olds are still using social media platforms covered by the ban, with 61% of young people who previously had accounts on these platforms still retaining access to at least one active account. Research from the University of Chicago finds 64.8% had used restricted social media platforms in the previous week and suggests that only 27% of children no longer hold accounts.[97]
- eSafety's own data suggests this proportion may be even higher, with surveys of parents suggesting that 69.4% of children still held a Snapchat account, 69.3% had access to a TikTok account and 69.1% an Instagram account.[98]
- Most under 16s who continue to hold accounts on one or more platforms covered by the ban hadn't needed to proactively circumvent restrictions, because in most cases platforms have failed to identify and remove their accounts in the first place. Among children who remained able to use accounts on restricted platforms, a significant majority (70%) said it had been 'easy' to circumvent the ban. More than three-fifths of children who continued to use YouTube (64%), Snapchat (61%), Instagram (60%) and TikTok (60%) said that that 'no action' had been taken by the platform to remove or deactivate an account they had before restrictions were introduced.
- Most children who used restricted platforms prior to the ban report that it has not improved how safe they feel online (51%). While three in ten (31%) reported they now felt safer, one in seven (14%) under 16s now reported feeling less safe online. eSafety data found that complaints of cyberbullying had increased 26% in January 2026 compared to January 2025, although we note this was only a few weeks after the ban came into force.[99]
- There are early indications that the ban may be having some positive effects for some children. Half of children (50%) who used restricted platforms prior to the ban coming into force report spending less time online. Among children who lost access to all their accounts, a majority self-reported positive impacts on their sleep, mental health and educational performance.
- However, when asked about the overall impact of the ban on their lives, around two-fifths (42%) of 12–15 year olds who used restricted platforms prior to the ban coming into force

[96] Molly Rose Foundation (2026) Australia's social media ban - is it working?
[97] Bursztyn, L, Duckworth, A et al (2026) Why Bans Fail: Tipping Points and Australia's social media man. University of Chicago, Becker Friedman Institute of Economics. Working paper No 2026-57
[98] Office of the eSafety Commissioner (2026) Social media minimum age: compliance update. eSafety surveyed 898 parents in January 2026.
[99] Pol, A (2026) No 'meaningful' shift from social media sites after Australia teen ban: govt report. Published in Tech Xplore, 30th April 2026

felt it had 'not had any impact'. Around a third felt it had a somewhat or very negative impacts (32%). Just over one in five under 16s (22%) felt the ban had a somewhat or very positive impact to date.

Early implications of Australia's implementation

The challenge of malign compliance

The early evidence from Australia raises substantial questions about the effectiveness of an outright social media ban. It also suggests that the arguments raised by proponents of an Australian-style ban - that it is a necessary firebreak to deliver swift and immediate improvements to children's safety – appear largely misplaced.

It is particularly striking that, despite this being a signature policy of the Albanese government, the primary reason for the ban's limited efficacy is widespread malign compliance by technology firms.

For example, among under 16s who retain access to TikTok accounts, three-fifths (60%) say that no action was taken by the platform to remove or deactivate a pre-existing account. One-quarter (25%) said they were successfully able to navigate the platform's age checks.[100]

During the consultation, MRF is aware that some pro-ban activists have claimed that the UK will be able to learn lessons from Australia's approach. However, we would raise significant doubts about whether UK regulators would be willing to adopt a fundamentally different approach to the design and enforcement of any access-based regulatory scheme than their counterparts in Australia.

Like much of civil society, MRF has been consistently critical of both Ofcom and the ICO for their reluctance to specify outcome- based metrics to determine the efficacy of Highly Effective Age Assurance, instead placing a set principle-based requirements on relevant providers.

Forthcoming research will show that measures to reduce exposure to harmful content, in part contingent on the use of HEAA, have been largely ineffective.

Separately, Ofcom has demonstrated exceptionally high levels of risk aversion in respect of its enforcement approach. As such, it is difficult to envisage the regulator aggressively and promptly pursuing large platforms for exercising a similar strategy of malign compliance to that which we are witnessing in Australia.

If the Government is minded to proceed with an Australian style ban, it must be able to demonstrate that it is willing to instruct legislation that can address the significant disincentives for technology companies to comply with the measures.

The Government must also set out what steps it intends to take to secure confidence that Ofcom will more assertively enforce these age-based restrictions than most other parts of the existing regulatory scheme. MRF notes that, at the time of writing, Ofcom has only opened one enforcement investigation against a presumed Category One platform (and this was only in response to the widespread public and political outcry against X over its rollout of nudification tools).

[100] Molly Rose Foundation (2026) Australia's social media ban – is it working?

Ofcom's first large-scale investigation, into a pro-suicide forum linked to at least 164 UK deaths, took almost 14 months before enforcement decisions were reached.

The scope of an Australian-style ban

Australia's social media ban has focused narrowly on ten social media platforms, with other social media platforms (such as Pinterest), gaming platforms (Roblox and Discord) and messaging services (such as WhatsApp and Telegram) left out of scope.

MRF's research suggests that, as a result of the ban, substantial proportions of 12 to 15-year-olds who had used restricted platforms prior to the ban coming into force report they are now using messaging platforms (39%) and gaming platforms (43%) more often.[101]

We are gravely concerned that an Australian-style social media ban risks harm migrating to other platforms out of scope, including gaming platforms such as Discord and Roblox. Gaming platforms form the leading edge of new and emerging online threats, including sadistic online exploitation by organised networks (Com groups).

In MRF and Resolver's recent threat assessment, gaming services were classified as 'critical risk' for Com networks establishing initial contact with their victims, while group servers and messaging platforms were deemed 'critical risk' in respect of grooming, radicalisation and harm and abuse.[102]

While it is entirely possible for the UK Government to establish the scope of any Australian-style ban differently, unless it is willing to adopt a relatively broad approach to the scope of restrictions – which in turn would raise substantial risks of unintended consequences and a range of child's rights concerns - there is an unacceptable risk that harm becomes misplaced rather than reduced.

We are aware that some pro-ban campaigners have advocated a 'ban plus' approach – that is to say, introducing restrictions on some platforms while strengthening regulation on others. MRF notes that the Government continues to resist calls to introduce a new Online Safety Act or to address substantial structural constraints associated with the current regime.

In the absence of further regulatory strengthening, the Government would effectively be shifting away from a safety-by-design approach to a predominantly access-based approach.

This will primarily have the effect of letting tech companies off the hook in terms of their safety by design and safeguarding responsibilities, despite the significant risk that children would remain freely able to use social media platforms at scale.

The evidentiary basis for introducing an Australian-style social media ban

Given the significant and unresolved question marks about the efficacy and implementation of Australia's approach, there is palpably no justifiable evidentiary basis to support proceeding with a fast-tracked Australian-style ban at this stage.

[101] Molly Rose Foundation (2026) Australia's social media ban – is it working?
[102] Resolver Trust and Safety, in partnership with Molly Rose Foundation (2026) Weaponised loneliness: Critical Harm Intelligence Briefing

That said, MRF encourages the government to continue to closely monitor the results from Australia's approach, including the outcomes from numerous long-term, systematic evaluations and academic reviews. This will enable the efficacy of regulatory measures to be determined, and where early data suggests that a ban may produce positive effects if it is capable of being scaled effectively, for the harm-benefit ratio to be properly assessed against other interventions, including compared to strengthened regulation that could deliver the same results.

We strongly encourage the Government to resist the temptation to 'move with the herd', simply because the loudest voices and diplomatic pushes from Australia and other countries may be encouraging it to do so. It is justifiably in the active self-interest of the Australian Government to build momentum for social media bans elsewhere, in the hope this will in turn boost compliance rates domestically.

We also caution the Government that it should recognise that the impacts of an Australian-style ban may be highly fluid – and that potential changes in usage, safety and online risk may not be unidirectional.

Experience from other countries suggests that the impacts of online access restrictions may denude or even reverse over time. For example, in 2011 South Korea responded to concerns about gaming addiction by introducing a ban on online gaming for children between midnight and 6am. Although the ban initially resulted in a reduction in time spent online, these improvements steadily eroded and within 4 years internet use had increased.

South Korea's government subsequently discontinued the policy, with evaluations finding 'practically insignificant effects' on time spent online, academic performance and adolescent sleep time. Studies showed that sleep time increased by an average of only 1.5 minutes per child, with internet addiction declining by only 0.7 percentage points over this period.[103] [104]

Research from the University of Chicago suggests that current compliance with the Australia social media ban (27%) is well-below the two-thirds threshold necessary to secure a shift in social norms.[105] The scale of this gap means that social norms are unlikely to shift enough to secure a long-term shift towards non-usage.

The study finds that compliance with the ban carries a 'significant social cost', with continued social media use being driven by significant social motives and strong network effects. Taken together, the lack of personal sanctions, the ease of circumvention, and the strong social costs of compliance are likely to prevent a 'high compliance equilibrium' from being reached and subsequently sustained.

The research concludes that current peer behaviour patterns means that compliance with the social media ban is 'more likely to diminish than to rise.' The study finds that four additional levers would likely be needed to either lift compliance above the threshold or lower the threshold itself:

[103] Lee, C (2017) Ex post valuation of illegalising juvenile online gaming after midnight: a case of shutdown policy in South Korea. Telematics and Informatics, 34(8)
[104] Choi, J et al (2018) Effect of the online game shutdown policy on Internet use, Internet addiction, and sleeping hours in Korean adolescents. Journal of Adolescent Health 62(5)
[105] Bursztyn, L., Duckworth, A. et al (2026) Why Bans Fail: Tipping Points and Australia's social media man. University of Chicago, Becker Friedman Institute of Economics, Working paper No 2026-57

1. The introduction of app caps on social media platforms. However, it is difficult to envisage how platforms could be instructed to ban users and at the same time also ration their time spent on online services.
2. Aligning the scope of the ban with school years rather than age. This would clearly represent a technically challenging undertaking and feasibility studies would therefore be appropriate.
3. Informational and marketing campaigns to encourage behavioural change and discourage social media use. Researchers suggest measures such as cash payments for verified non-use, vouchers for activities that fill the freed time, or even tax credits for parents.
4. Investment in in-person coordination infrastructure, after schools' clubs and routines and other 'third space' provision. During MRF's 'open space' event, young people repeatedly raised the cost and access barriers to in-person activities and cited their online usage as in part driven by a lack of physical and in person alternatives.

Potential unintended consequences of an Australian-style social media ban

Unintended consequences for wellbeing and youth mental health

While bans may reduce exposure to harmful content in the short-term, evidence shows many young people rely on social media for connection, identity exploration and support. As set out in Section 1, For LGBTQ or neurodiverse children, being online can offer huge benefits around identity, self-esteem and peer-support.

Early evidence from Australia suggests that, despite the limited efficacy of the ban so far, adverse impacts among potentially vulnerable young people are being felt.

Australia's crisis service for young people, Kids Helpline, reports that young people have sought specific support around dealing with the impacts of the ban, including young people experiencing high levels of distress and suicide ideation. Highest levels of emotional distress were recorded among younger children, girls and children who are neurodivergent. Contacts reported feeling cut off from support networks, including in one case a young person who had used social media as a support tool to manage self-harm urges.[106]

Separately, one in ten 12-15 year olds accessing Australia's youth mental health service, Headspace, reported that the social media ban has been a factor in them seeking immediate help. While clearly highly preliminary data, Headspace reports girls and young people identifying as LGBTQ+ were more likely to seek mental health support.[107]

Displacement effects for young people experiencing suicide and self-harm ideation

While social media is a deeply imperfect means for children and young people to access peer support, MRF is concerned that an outright social media ban could deprive young people

[106] Rintoul, C (2026) 'Distressed' teens turn to Kids Helpline following social media ban, saying they've lost support networks. Published in The West Australian, 7th January 2026
[107] Wilson, C (2025) One in 10 these seeking mental health support from headspace site social media ban as an issue. Published on Crikey.com.au16/01.26, 16th January 2026

experiencing suicide and self-harm ideation and/or behaviours of important sources of connection and peer support.

While the current design and operation of social media results in a mix of harmful and protective effects (and at present harmful effects predominate), social media provides an important source of peer support; children are able to benefit from assortative relating effects; and studies have found evidence of self-harm mitigation and reduction strategies, the promotion of self-harm recovery, and a range of emotional, cognitive and physiological impacts that mitigate or reduce self-harm urges and acts.[108]

Our analysis of pro-suicide forums finds that many young people join these spaces seeking help and support. This includes examples of young people posting about their negative experiences of mental health support services, a lack of peer support and a sense that they have tried all other avenues open to them.[109]

MRF is highly concerned that a reduction in comparatively lower risk sources of peer support, including social media, could shorten the pathways before young people find themselves exposed to very high-risk pro-suicide forums – and that as a result, a cohort of highly vulnerable younger teens and adolescents could end up being exposed to very significant high-risk environments.

Barriers to disclosing exposure to online harm

Molly Rose Foundation has strong concerns that any ban will introduce barriers to young people disclosing online harm out of fear that they will be blamed, punished, or have their devices confiscated for violating age restrictions. There are clear parallels for this in other safeguarding contexts, with fear of negative consequences preventing young people from speaking up.

The National Crime Agency and National Police Chief's Council highlighted this as a key risk of a blanket ban in the context of online offences against children. A child who has been coerced into sharing intimate images, for example, 'should receive help, not consequences for bypassing age checks.' [110]

A cliff edge for older teens and a sharp increase in risks while a ban takes effect

Any ban would introduce a deeply damaging cliff edge for older teens – and particularly girls - when they are suddenly exposed to poorly regulated online spaces on their sixteenth birthday. Girls may face an immediate barrage of harms, from misogyny to sexual threats, self-harm and eating disorder content, while being wholly unprepared to safely manage them.

Any announcement of a ban would also have an immediate chilling effect on years of hard-earned progress on safety by design. Preparations for a ban would likely drain energy and resource away from ongoing efforts to improve protections for children, while in-scope platforms would be disincentivised to cooperate with regulation or innovate in responsible design, leaving children at greater risk in the potentially lengthy period before implementation.

[108] Susi, K. et al (2023) Research Review: Viewing self-harm images on the Internet and social media platforms: systematic review of the impact and associated psychological mechanisms. Journal of Child Psychology and Psychiatry, 64(8). pp1115–1139
[109] Molly Rose Foundation (2025) Missed Chances, lost lives
[110] National Police Chief's Council and National Crime Agency (2026) Under-16s access to social media – NPCC and NCA's response

Damaging the ability for teens to gain critical online and algorithmic literacy skills – harming their long term life chances

Bans risk undermining the next generation's preparedness for adulthood, and particularly giving young people the critical online literacy and algorithmic literacy skills that will allow them to stay safe as teens, thrive as adults, and that will underpin our future digital and AI economy. It is far more difficult to develop these crucial skills in schools if we are teaching children about the risks of something they are no longer supposed to access.

There is also clear evidence that social media can be a powerful tool for learning, expression, creativity and the fulfilment of other children's rights – benefits which we can build on through effective regulation.

As such, MRF considers the social and economic impacts of a ban to be regressive.

Views and preferences of parents

Parents are profoundly concerned about children's online safety and overwhelmingly support decisive further action. However, it appears that most parents remain focused on the ends, rather than a particular set of means.

MRF data shows that three quarters (73%) of UK adults support new legislation to strengthen regulation and better protect children and young people from online harm, with support for regulation consistently polling more strongly than an Australian-style ban.[111] Recent polling for The Mirror found support for a ban at 66%.[112]

Parents attach a clear 'confidence premium' to the Government's next steps – demanding decisive action but wanting confidence that the measures taken will actually work. Our polling shows that equal proportions of UK adults support the UK Government considering the views to a great or moderate extent both the views of parents (82%), but also the evidence around social media and its impacts (82%.)

Two-thirds of adults (69%) support children's views being substantially taken into account.[113]

Polling from Public First reinforces the sense that support for bans is relatively superficial – although 64% support an Australia-style ban, 68% say it would not work, and 50% of parents say they would still allow access to social media access if a ban was in place.[114]

Strikingly, when asked whether they would prefer a social media ban against a range of alternative policy options, UK adults consistently opt for alternative measures. These measures include mandatory child-safe versions of apps (preferred by 53% to 39% in favour of bans), stronger parental controls (55% to 39%) and improved media literacy education in schools (50% to 42%.)[115]

[111] Molly Rose Foundation (2026) Online Safety Consultation: the next steps that adults and parents want
[112] Huskisson, S. (2026) Mirror poll reveals even more people support social media ban for under 16s – see results. Published in the Mirror, 7th May 2026
[113] Molly Rose Foundation (2026) Online Safety Consultation: the next steps that adults and parents want
[114] Tech at Public First (2026) To Ban or not to Ban
[115] Ibid.

MRF encourages the Government to carefully reflect upon public and parental views when deciding its next steps. Despite the voluble campaigns supporting a social media ban, polling consistently demonstrates that most adults would support – and in reality prefer – alternatives to a social media ban.

In this respect, parents are broadly aligned with the consensus of civil society, child safety and academic opinion, with a clear preference for decisive action that builds on, but does not replace, regulatory and evidence-led approaches.

Views and preferences of children

Children and young people similarly prefer other alternatives to a blanket Australian-style ban.

Recent polling of 70,000 UK schoolchildren found that an overwhelming proportion of young people opposed a social media ban, including 89% of secondary school pupils.[116]

However, this should not be taken as evidence that children do not support further action. Young people themselves recognise that the status quo is entirely unacceptable, with clear support for stronger protections that can protect and promote their safety and wellbeing online.

Recent polling by Girlguiding, for example, found that nearly three-quarters (72%) of 10-16 believed that social media companies should be doing more to protect young people online. However, in line with early findings from Australia, only 15% felt that a ban would make them feel safer online, with children preferring action to address the drivers of harm. Two thirds (69%) of 10-16 year olds say they would prefer to know if a platform is safe than be banned from it.[117]

What should happen: risk-based age ratings underpinned by an enforceable minimum age of 13

MRF strongly supports the introduction of an enforceable minimum age for social media, messaging services, and high risk gaming platforms at age 13. This move will ensure that platforms finally start to enforce their minimum age rules effectively – and will bring the legal and regulatory requirements on relevant user-to-user services in line with public expectations.

An enforceable minimum joining age should also be accompanied by the introduction of new risk-based age ratings based on a platform's functionalities. These would see Ofcom determine minimum age limits for a range of functionalities, with higher risk functionalities attracting higher minimum age ratings.

By implication, this regulatory approach would finally incentivise innovation in safety- and age-appropriateness by design – with platforms incentivised to develop lower risk and more age-appropriate versions of their products if they wish to continue for them to be available to teens.

Ofcom should be required to establish an updated definition of Highly Effective Age Assurance, setting out clear and measurable metrics against which the efficacy of a platform's age assurance mechanisms can be meaningfully judged.

[116] Results of in-school polling undertaken in PSHE lessons. Conducted by Votes for Schools in January 2026

[117] Girlguiding (2026) Social media companies need to do more to protect young people

The regulator should face a new legal duty to produce an annual report setting out whether and to what extent young people are able to circumvent age assurance measures, and more broadly the efficacy of any measures in its Codes which are effectively contingent on the robustness of the HEAA arrangements in place.

Section 3: High risk features and functionalities

Summary

Molly Rose Foundation strongly supports proposals to ban high-risk functionalities for children under 16.

As it stands, social media platforms have been characterised by the rollout of high-risk functionalities and design features, with safeguarding and/or product safety measures often being retrofitted only in response to high profile tragedies or civil society, media and political pressure.

It is also clear that regulation has achieved insufficient progress to embed a safety-by-design culture in regulated firms. The regulatory regimes overseen by both Ofcom and the ICO have wholly failed to shift the strategic and design incentives on user-to-user platforms.

We believe this is primarily because regulation has largely been constructed around whether it is proportionate to take steps mitigate risks, rather than an approach that sets out clear outcome-based sets of requirements (this would require platforms to demonstrate that products and/or functionalities are fundamentally safe before they are rolled out widely.)

In this chapter, we set out our views on high-risk functionalities that directly facilitate acute and cumulative harms. We propose:

- A conditional ban on personalised recommender systems unless platforms meet strict safety and wellbeing conditions, including the proactive recommendation of high quality content
- A conditional ban on AI chatbots unless they meet strict safety and wellbeing conditions
- Restrictions on livestreaming, disappearing messages and sending nude images

Restrictions on these features should form part of a broader regime of risk-based age ratings based on a platform's functionalities, with Government and the regulator determining age limits for a broader range of functionalities than those listed here. They should also apply where relevant across the stack.

In the following chapter, we will set out our position in respect of a broader range of addictive and engagement-based design features, linked to acute but also a broader range of chronic harms, for example infinite scroll and autoplay.

This section relates to consultation questions 12-15, 21-25 and 26-30

It also includes MRF's response to questions 49-51 on high quality online content

Personalised recommender systems

MRF strongly supports a conditional ban on personalised recommender systems, with platforms being required to meet a series of strict conditions if they are to continue to offer them.

MRF's research has shown that recommender systems remain the highest risk functionality for exposure to harmful content. Among 13–17-year-olds who had been exposed to suicide content in the previous week, 57% said they had seen this on algorithmic feeds.[118]

If platforms are to continue to offer personalised feeds, this must be under a set of five strict conditions, set out below:

1. Platforms must meet clear outcome-based measures that prevent exposure to harmful content

Social media services must face clear and unambiguous outcome-based requirements to prevent harmful content, including Primary Priority and Priority Content, from being discoverable or made accessible to under 16s.

The current regulatory requirements have largely failed to prevent harmful content from being algorithmically recommended. This is largely because of how Ofcom has developed its Code of Practice requirements, with platforms being required to prevent exposure to harmful content only where there has already been identified (effectively operationalising a safety-by-design measure through a content moderation lens.)

2. Platforms must offer under 16s meaningful agency over their content recommendations

Under 16s must be given a comprehensive range of tools to exercise agency over their recommender feeds, including the opportunity to reset their feed, turn-off personalised recommendations, and to suggest a list of topics which cannot be personally recommended to them.

Teenagers should also have the opportunity to provide meaningful feedback on content being algorithmically recommended to them, with personalised feedback being appropriately weighted in respect of future content recommendations.

While Ofcom has introduced some welcome measures in this respect in its Children's Code of Practice, in reality these provisions appear to have delivered only a modest effect, with MRF analysis finding that both Instagram and TikTok had actively gamed these requirements - with young people able to offer both positive and negative feedback on the content they were being shown.[119]

In turn, this meant that young people could be served additional amounts of similar harmful material, including posts referencing suicide ideation, suicide behaviours, and content that reinforced and normalised intense feelings of hopelessness, misery and despair.

3. Platforms must meet diversity-by-design requirements

[118] Molly Rose Foundation (2025) Children's exposure to suicide, self-harm, depression and eating disorder content on social media
[119] Molly Rose Foundation (2025) Pervasive by Design

Social media platforms should face strict content plurality requirements, with algorithms set to healthy defaults that promote a variety and balance of topics, rather than enabling young people to be algorithmically bombarded with discrete or interrelated topics that can exacerbate the risk of cumulative harm.

MRF recommends that personalised algorithms face caps on the total number of posts that can be shown in respect of any one topic or set of related themes. For example, feeds could be restricted so that no more than 10 per cent of content recommendations relate to any one theme or interest.

4. A ban on personalised recommendations through notifications, update feeds and similar means

As explored in more detail in Section 4, platforms should be prohibited from using alerts, push notifications and emails to suggest personalised content recommendations to under 16s. Restrictions should also be introduced on content recommendations appearing in update feeds, commonly used on platforms such as Instagram, TikTok and Pinterest.

MRF research has consistently shown how Instagram continues to use off-platform user engagement techniques, including personalised content recommendation emails and app notifications, to nudge teenage users back onto the site.[120]

5. Proactively recommended trusted, high quality sources of content

As a pre-condition for offering personalised recommender feeds to teens, platforms should also be mandated to give due prominence to high quality content from a range of trusted providers, including trusted sources of mental health and well-being support, education providers, and the U.K.'s public service broadcasters.

The Government should build on the 'must carry' duties already in place for PSBs under the Media Act, introducing a clear process for defining 'trusted' providers. Ofcom should also be tasked with developing and regularly updating a set of high-quality content principles.

Ofcom should specify a minimum proportion of high-quality content that must be recommended to under 16s accounts, which we recommend is set at no less than 20% of all content recommended. This reclaiming of algorithmic feeds would offer clear and immediate benefits to children's lives - promoting their well-being; exposing them to a broader range of high quality, age-appropriate content; and delivering wider societal benefits such as the fostering of British values, including respect, tolerance, equality, and an appreciation of fundamental rights.

The consultation's proposals in respect of wholly voluntary measures are entirely unsatisfactory in this regard.

AI chatbots

MRF is deeply concerned about the risk profile of AI chatbots for teens, particularly chatbots with anthropomorphic characteristics.

[120] Ibid.

Until the government commits to substantive improvements to the risk assessment and mitigation duties in the Online Safety Act, including a requirement to introduce robust product safety testing requirements, we therefore support restrictions on AI chatbots for under 16s.

MRF envisages these restrictions taking the form of a conditional ban, with AI companies being prevented from offering services to children in their current high-risk form. However, we recognise that AI chatbots can offer a range of potential benefits to children, particularly if developed in a safe and age-appropriate way.

We would therefore support platforms being able to offer safer and more age-appropriate versions of their products, developed in the form of a 'walled garden' proposition. MRF envisages chatbots having to demonstrably meet a range of safety and wellbeing-by-design criteria, including but not limited to:

- the removal of anthropomorphic characteristics;
- design considerations that minimise the risk of unhealthy parasocial attachments, for example through the ability to mimic friendships, relationships and empathy;
- design considerations that prevent the risk of adverse physical or mental health outcomes, including suicide and self-harm risks. This should include appropriate product safety testing and safeguarding checks, the ability to recall interactions across sessions, and strict criteria for safeguarding referrals;
- the elimination of engagement-based design features, including features that encourage further questions and interactions;
- restrictions on the ability of AI chatbots to provide responses on sensitive topics, including mental health, self-harm and suicide ideation, with exemptions in place for clearly defined therapeutic use cases;
- the ability for AI discussions to be retained and accessible for safety and investigatory purposes. Meta recently rolled out a new 'Incognito mode' across WhatsApp, Facebook and Instagram, an unacceptably high risk decision which will effectively prevent Meta, law enforcement and coronial proceedings from being able to access AI records in the event of a child's death.

Disappearing messages

MRF would support restrictions on disappearing messages for under 16s, either in the form of an outright ban and/or the introduction of 'cooling off' periods that would prevent disappearing messages being sent between mutual followers for the first month after a connection is accepted.

In May 2026, Instagram launched a new ephemeral message feature called 'Instants'. This feature allows targeted sharing of photos and videos, with the photo or video only viewable once and disappearing entirely after 24 hours. Described by Wired as a 'Snapchat clone for thirst traps',[121] there are obvious risks that this feature could be used in a range of harmful and

[121] Rogers, R (2026) Instagram's new Instants app is a Snapchat clone for thirst traps. Published on Wired, 13th May 2026

potentially damaging ways, including for the purposes of bullying, sending age-inappropriate content, and to instruct and encourage harmful behaviour.

Livestreaming

MRF would support restrictions on livestreaming functionalities for under 16s, given the marked risks they currently pose to children's safety.

Livestreaming functionalities can play a key role in suicide and self-harm offences, with evidence that they are used by perpetrators to groom, coerce or encourage children into acts of self-harm and suicide. MRF are particularly concerned by their role in the threat profile posed by Com Networks, with a recent threat assessment rating livestreaming functionalities as 'critical risk' for use by perpetrators in both grooming victims and the perpetration of harm and abuse.[122] An FBI advisory described both how prospective group members are required to live-stream videos of themselves committing violent offences in order to be granted entry to a group, and how groups use messaging platforms to threaten, manipulate or coerce vulnerable minors 'into recording or livestreaming self-harm, sexually explicit acts, and/or suicide'.[123]

Evidence also shows that livestreaming can the increase the risk of children being exposed to suicide and self-harm content.[124]

Action to restrict these functionalities would be particularly welcome given the inadequacy of existing measures to protect children. Proposals currently being consulted on by the regulator, for example, are targeted solely at once a victim is already on a livestream, rather than aiming to intervene upstream, introduce friction and disrupt well-established pathways to harm that end on livestreams.[125]

Sending and receiving images and videos containing nudity

Molly Rose Foundation supports action to strengthen measures to prevent children receiving and sending nude images. This is important not only to prevent child sexual abuse, but to prevent children being exposed to a broader web of group-based online threats, including the threat posed by Com Networks.

This must happen at the platform level, with strict requirements on regulated services to detect and block the transmission of illegal content.

Ultimately, Government should also take action at the device level, legislating to require the default installation of nudity detection measures that prevent nude images being created, shared, or viewed on handsets and other devices registered to under 18s.

[122] Resolver Trust and Safety, in partnership with Molly Rose Foundation (2026) Weaponised loneliness: Critical Harm Intelligence Briefing
[123] Federal Bureau of Investigation (2023) Violent Online Groups Extort Minors to Self-Harm and Produce Child Sexual Abuse Material
[124] Ofcom (2024) Children's Register of Risks
[125] Molly Rose Foundation (2025) Molly Rose Foundation response to Ofcom's consultation on Additional Safety Measures

Section 4: Persuasive design features

Summary

Molly Rose Foundation strongly supports proposals to restrict addictive and engagement-based design features for under 16s.

Persuasive design features are linked to a range of chronic harms, including the opportunity costs of excess screentime, compulsive behaviour, damaged relationships, and other negative consequences for health and wellbeing. They are also linked to more acute harms – with persuasive design features encouraging young people to spend excess time on high risk product surfaces, or directly driving exposure to harmful content.

We also know that parents have profound concerns about the impact of these features. Over half (55%) of parents of 8-17 year olds say they think their child's screen time is too high, with higher rates among parents of older children.[126]

As it stands, however, these harms are poorly targeted by the current scope and operation of the Online Safety Act, which focuses on acute risks, such as illegal content and activity and content that it is harmful to children.

Action to tackle persuasive design features would be a clear signal of intent that Government is willing to go after the engagement-based business models that continue to prioritise profit over children's safety and wellbeing.

As set out in Section 1, legal disclosures relating to Meta's Project Daisy and recent US trials confirm that platforms deliberately deploy addictive and high-risk design features in order to maximise engagement and therefore revenues, despite internal evidence of their negative impact on children's wellbeing.[127] [128]

In this section, we set out our views on persuasive design features that directly facilitate both chronic and acute harms. We propose:

- Restrictions on autoplay and infinite scroll for under 16s
- Restrictions on affirmation features for under 16s, including 'streaks'
- Restrictions on notifications and other prompts for under 16s, other than those associated with messages from close friends and family
- Restrictions on a broader range of persuasive design features and dark patterns for under 16s

As above, restrictions on these features should form part of a broader regime of risk-based age ratings based on a platform's functionalities, with Government and the regulator determining age limits for a broader range of functionalities than those listed here. They should also apply

[126] Ofcom (2026) Children and Parents: Media Use and Attitudes Report
[127] Mirza, R (2023) Case Study on Online Youth Harms – Project Daisy
[128] Hays, K. (2026) Campaigners welcome Meta and YouTube's defeat in landmark social media addiction trial. Published in BBC News, March 2026

where relevant across the stack – encompassing social media, gaming platforms and messaging services.

Ultimately, action to restrict persuasive design features within the scope of this consultation should be a downpayment on new legislation focused on delivering wellbeing by design, on which more detail is provided in section 7.

This section relates to consultation questions 16-18.

Infinite scrolling and autoplay

MRF supports proposals to restrict both infinite scrolling and autoplay for under 16s.

As set out in section 1, in combination with recommender systems, these features are deeply embedded within engagement-based business models, deliberately creating a seamless and personalised experiences that makes it difficult for users – and particularly children – to regulate the time they spend on platforms.[129] [130] As such, they undermine children's right to agency over their online experiences, and play a key role in driving compulsive and excess use of online platforms, which has been linked to a range of chronic harms.

Given that they incentivise high levels of use of recommender feeds, they also drive exposure to harmful content, with MRF research consistently identifying recommender feeds as the single biggest vector for children's exposure to suicide, self-harm and intense depression content.[131] [132]

Affirmation features

MRF supports proposals to restrict affirmation features for under 16s. This should include any visible quantification of engagement, including like counts, follower/friend counts and scores (e.g. 'Snap scores').

As set out in Section 1, affirmation features are a key aspect of engagement-based business models, encouraging high levels of engagement with platforms by exploiting adolescents' heightened desire for social affirmation and comparison. They are also linked to negative mental health impacts, with increased sensitivity to likes predicting declines in mental health over time,[133] and trials to remove these features leading to improvements in children's wellbeing.[134]

[129] Behavioural Insights Team (2025) Behavioural Audit of Online Services: Key findings report
[130] American Psychological Association (2026) Potential Risks of Content, Features, and Functions: A closer look at the science behind how social media affects youth.
[131] Molly Rose Foundation (2025) Children's exposure to suicide, self-harm, depression and eating disorder content
[132] Molly Rose Foundation (2025) Pervasive by Design
[133] Ibid.
[134] Mirza, R (2023) Case Study on Online Youth Harms – Project Daisy

Affirmation features can also contribute to more acute risks, with likes and reposts on suicide and self-harm content shown to provide validation and reinforce negative thought patterns and behaviours.[135]

We also support restrictions on **streaks** for under 16s. Streaks can exert extreme social pressure on young people, and are associated with elevated levels of stress, problematic smartphone use and other negative impacts.[136]

Notifications and other prompts

MRF would support proposals to restrict engagement-based notifications and other prompts for under 16s, including notifications associated with affirmation features, and those associated with personalised content recommendations. These restrictions should apply across all surfaces where these are deployed, including push notifications sent to a user's devices, on-platform notifications, and email notifications.

As set out in Section 1, engagement-based notifications can have a corrosive effect on children's wellbeing, particularly when they are frequent and intrusive. This includes undermining children's agency, disrupting focus, and contributing to compulsive use of online platforms. They can also drive exposure to more acute risks, with evidence that platforms continue to use dark patterns and scarcity principles to prompt teen users to log onto to platforms and view potentially harmful content.[137] As such, there is no clear case for under-16s to have access to affirmation and recommendation notifications that primarily seek to maximise engagement over children's safety and wellbeing.

However, children do benefit from positive social connection online, with notifications helping to keep them connected to friends and loved ones. In this context, we support allowing messaging and call notifications from close friends and family.

However, this exemption must still meet strict wellbeing criteria, including ensuring users have meaningful choice over the notifications they receive. This must include easily accessible tools to mute or reduce notification frequency, at a reasonable level of granularity (e.g. rather than a 'all or nothing' approach).

Other persuasive design features and dark patterns

MRF supports action to restrict a broader range of persuasive design features and dark patterns.

We encourage Government to do this within the scope of this consultation by including a broader range of features in scope of age-based restrictions, which should then be operationalised via a risk-based age rating regime. Examples of features potentially in scope can

[135] Susi, K et al (2023) Research Review: Viewing self-harm images on the Internet and social media platforms: systematic review of the impact and associated psychological mechanisms. Journal of Child Psychology and Psychiatry, 64(8). Pp1115–1139
[136] Essen, C & Ouytsel, J (2023) Snapchat streaks – How are these forms of gamified interactions associated with problematic smartphone use and fear of missing out among early adolescents? Telematics and Informatics Reports, 11
[137] Molly Rose Foundation (2025) Pervasive by Design

be found in the 5Rights Foundation's 'Disrupted Childhood'[138] and Ofcom's report on Persuasive Design and child financial harms.[139]

However, ultimately this should be delivered as part of broader action to expand the scope of the Online Safety Act to encompass wellbeing-by-design, as described in Section 7. If a platform wishes to offer an engagement-based feature or functionality, they must be able to demonstrate not only that this has been robustly product tested and risk assessed as safe, but that it is consistent with the best interests of the child and does not nudge or manipulate users to make choices or take actions that they might not otherwise choose to do.

As part of this, there must be a blanket prohibition on dark patterns for all users – features that don't just 'persuade' but manipulate, pressure or mislead users into making decisions that reduce their safety, privacy or wellbeing.

[138] 5Rights Foundation (2023) Disrupted Childhood: The cost of persuasive design
[139] Discovery Research on behalf of Ofcom (2025) Persuasive design features and potential child financial harms

Section 5: Digital and media literacy

Summary

Molly Rose Foundation strongly supports proposals to provide additional support to children and families around digital and media literacy.

In MRF's view, digital and media literacy is a fundamental life skill that can inoculate today's young people from harm, while preparing them to thrive in adulthood.

In conjunction with effective regulation, digital and media literacy can help turn the tide on preventable online harm. As technology moves at pace, children's safety and wellbeing depends on their ability to recognise online risks, behave responsibly, critically evaluate content and interactions, and take practical steps to manage their experiences.

Digital and media literacy can also provide young people with the tools to flourish in adulthood – giving them the critical thinking tools to thrive in an AI and digital economy, to prepare for a new voting age of 16, and to deal with the increasing threats a fractured information ecosystem poses to our democracy. Information literacy, for example, is now key to workforce productivity,[140] while the House of Lords report on media literacy concluded that it is central to empowering 'informed and responsible citizens.'[141]

In this context, Government must first recognise digital and media literacy as a foundational skill in the modern era, with this reflected in a clear strategic commitment and ownership at the highest levels of Government.

This must be followed by measures to strengthen media literacy across the stack. Ultimately, Government must revisit the scope of this consultation and recognise that further legislation is required to mandate action on media literacy by platforms, unlock funding for the sector, and support teachers and schools. Key actions must be:

- Mandating media literacy by design as part of a broader wellbeing-by-design duties on platforms
- Supporting schools to embed a cross-curricular approach to digital and media literacy, and training every teacher for the digital age
- Delivering sustained funding through a polluter pays approach

We also welcome the Government's willingness to explore which areas of digital and media literacy children and families most need additional help with. As set out below, we strongly encourage the Government to offer additional support around platform and algorithmic literacy.

This section relates to consultation questions 44-47.

[140] National Foundation for Educational Research (2023) The Skills Imperative 2035: An analysis of the demand for skills in the labour market in 2035

[141] House of Lords Communications and Digital Committee (2025) Media Literacy

Placing strategic emphasis on digital and media literacy

While the prize of high quality digital and media literacy is considerable, the resulting safety, social and economic benefits will only be unlocked if all parts of Government – including the Treasury and No10 – recognise the immense strategic significance of delivering a step-change in how digital and media literacy is supported and delivered.

As it stands, although the Government has announced its Media Literacy Action Plan (2026-2029) it seems clear that the broader strategic social and economic value of high-quality digital education – and the return on investment that it brings – remains unrecognised. While a welcome signal of intent and cross-Government action, under the plan media literacy still has no ownership at the highest levels of Government, there is limited new funding, no metrics against which to assess the success of the plan against, and little detail on how different departments and Ofcom will coordinate work in line with their separate strategies.

MRF strongly recommends that the Government prioritises an ambitious, well-funded and cross-cutting strategy that is owned at the cabinet level. Action on digital and media literacy must be commensurate to its status as a core foundational life skill, and as an investment in life chances, brain capital and our economic potential.

Clear requirements around media literacy by design

MRF strongly encourages the Government to introduce robust requirements on platforms to deliver media literacy by design. New legislation should put media literacy by design on a statutory footing, imposing clear duties on platforms as part of wider action around wellbeing-by-design set out in Section 7.

Platforms have a vital role to play in supporting children to think critically and have meaningful agency over the their online experiences. This should include empowering children to think critically about both content and the design of the digital environment (including the implications of algorithmic feeds), providing transparent information to support informed choice, supporting meaningful and continuous agency over online experiences, and not deploying persuasive design features and dark patterns that frustrate media literacy.

As it stands, however, the vast majority of requirements are voluntary, including recommendations currently being consulted on by Ofcom.[142] This approach is unlikely to deliver meaningful change, particularly as only a handful of platforms have pledged to deliver Ofcom's similarly voluntary Best Practice Design Principles for Media Literacy.

If Government is unable to put media literacy by design on a statutory footing, we strongly encourage it to work with the regulator to explore how existing powers can be used to drive action. This should include the introduction of far more stretching and outcome-focused guidance than is currently being consulted on, alongside the use of transparency and information gathering powers to drive compliance. As a minimum the regulator should adopt principles from its approach to for driving compliance with VAWG guidance, including close supervision arrangements with platforms and regular public reporting on progress.

Securing sustained funding through a polluter pays approach

[142] Ofcom (2025) How to promote Media Literacy

MRF proposes that the Government look to amend the OSA to ringfence funds from regulatory enforcement action to support education and prevention initiatives via a centrally allocated pot. This approach draws directly on the 'polluter pays' principle, enables improved educational outcomes while being exchequer neutral, and builds upon the working precedents in other regulated markets.[143]

As it stands, funding for digital and media literacy initiatives has long been characterised by short-termism and instability, leading to inefficiencies, duplication, limited reach, and failures to adequately scale or evaluate initiatives.[144] Over-reliance on services' contributions also establishes an unhealthy dependence on platforms who may not have it in their best interests to develop digital literacy skills that may affect their business model (i.e. platform and algorithmic literacy), or who are not able to deliver effectively-targeted or high-quality interventions.

Supporting schools to embed a cross-curricular approach to critical digital and media literacy, and training teachers for the digital age

Schools are the crucial site of intervention for digital and media literacy. Digital and media literacy needs to be strongly embedded across the entire curriculum and all primary and secondary age groups. It must also be delivered by a confident, well-trained workforce who understand their role as part of a school-wide approach.

In England, there are good foundations to build on, including commitments to ensuring digital and media literacy are 'embedded' into the revised national curriculum, to making subject-specific changes in Citizenship, Computing and English, and to expanding RSHE guidance.[145]

In this context, the Government must introduce new guidance to drive best practice and ensure the whole curriculum sings in harmony. This must set clear expectations for how digital and media literacy should be embedded into education for all ages, and should include:

- A consolidated framework for how all subjects and wider activities should 'sing in harmony' to build digital and media literacy throughout every key stage, while avoiding duplication.
- Evidence-based best practice for teaching digital and media literacy, including core principles like platform and algorithmic literacy, and clear outcomes to assess progress.
- Guidance that is applicable to children with SEND and who are outside of mainstream education, including those in alternative education, pupil referral units or SEN schools.

Every teacher must also be trained to take responsibility for supporting children's digital and media literacy, with many currently lacking appropriate training or the confidence, knowledge or skills to address certain topics.[146] [147] MRF encourages the Government to explore opportunities to embed digital and media literacy in Initial Teacher Training or Newly Qualified Teacher training, with enhanced provision for key subjects. This should build on the online safety

[143] Ofgem's similar Energy Redress Scheme has led to £150 million of additional for measures to support vulnerable consumers.
[144] DSIT (2023) Cross-sectoral challenges to media literacy
[145] UK Government (2025) Government response to the Curriculum and Assessment Review
[146] Molly Rose Foundation and University of Bristol survey of 131 secondary school staff, not yet released
[147] Internet Matters (2023) Data Briefing: online safety in schools.

elements of existing safeguarding training and prioritise broader foundational skills – including platform and algorithmic literacy – as well as best practice. This must be supported by ongoing work by the Department for Education to commission new CPD modules and support the development of a consolidated and regularly updated bank of high-quality resources.

Assessing the quality of digital and media literacy education is essential to drive improved outcomes and support. However, as it stands, children's attainment is poorly understood. Education departments and inspectorates across all nations should therefore ensure inspection frameworks assess a comprehensive range of digital and media literacy outcomes, rather than solely in a safeguarding or personal wellbeing context. Like teachers, inspectors must be trained to confidently identify and promote best practice.

Platform and algorithmic literacy

MRF encourages the Government to offer additional support to children and families around platform and algorithmic literacy.

Platform literacy refers to critical engagement with the digital environment itself, including understanding the central role of algorithms and data-drive decision-making in shaping what we see, how persuasive design manipulated our online choices and behaviour, and the pervasive influence of business models and time-spent commercial metrics.[148] [149] Critical algorithmic literacy is particularly important. This means understanding what algorithms are and how they work, being able to critically engage with their effects, and having the skills to shape or influence these systems. Improving algorithmic literacy not only protects young people from the mental health risks posed by recommender systems, it helps to protect them from content-based harm of all kinds and make the most of how information works in the modern world.[150] [151]

Through gaining these cross-cutting skills, children move from passive consumers to active participants, better able to manage the numerous ways digital design puts them at risk, and to develop healthy and more deliberate relationships with digital products.

Despite these benefits, forthcoming MRF research shows marked gaps in provision and confidence around algorithmic literacy in English secondary schools, and therefore missed opportunities ranging from suicide prevention to preparing children for future workplaces.[152]

In this context, MRF encourages the Government to place additional emphasis on platform and algorithm literacy support for children and families across a range of contexts, whether in community settings, schools, campaigns or education programmes. In England, for example, there is a clear opportunity to place renewed emphasis on these competencies in programmes of study being developed following the Curriculum Review in England, commissioning new resources, and embedding them into strengthened cross-curricular guidance and training.

[148] See Polizzi, G. (2020) Why we need critical digital literacy to participate in democracy
[149] Livingstone, S. et al (2025) Can platform literacy protect vulnerable young people against the risky affordances of social media platforms? *Information, Communication & Society* 29 (2)
[150] Regehr, K et al (2025) Normalizing toxicity: the role of recommender algorithms for young people's mental health and social wellbeing. Frontiers in Psychology, 16, 14, Article 1523649
[151] Winstone, L. (2024). *Developing algorithmic literacy for positive social media engagement*. C. Fellowship.
[152] Molly Rose Foundation and University of Bristol survey of 131 secondary school staff, not yet released

Section 6: Other Interventions

In this section, MRF responds to other proposals set out in the consultation.

We set out our position on:

- Why raising the age of digital consent should not be a priority
- Support for time limits on children's access to services that meet rigorous design standards
- Support for parental controls that meet rigorous design standards
- VPN restrictions

This section relates to consultation questions 8-11, 19-20, 36-40, and 52-54

Age of Digital Consent

MRF does not support the prioritisation of raising the age of digital consent as an intervention to improve children's safety and wellbeing online.

While proposals to raise the age of digital consent may offer potential benefits, in practice we expect these would be limited.

In particular, MRF has strong concerns about whether any change to the age of digital consent would be effectively enforced. As it stands, the Information Commissioner's Office (ICO) have been very reluctant to robustly enforce either the digital age of consent or the broader Age Appropriate Design Code – something made clear by the seven-in-ten 3-12 year olds who have an account on at least on social media, video, messaging or livestreaming site.[153] As set out in Section 2, this is in part due to the regulator's failure to specify outcome-based metrics to determine the efficacy of Highly Effective Age Assurance. Until this meaningfully changes, raising the digital age of consent will have little impact.

MRF also has broader concerns around the inefficiency and uncertainty that results from Ofcom and the ICO sharing responsibility for age assurance. Even assuming effective collaboration under the Digital Regulation Cooperation Forum, both regulators have recognised the need for more clarity on the interaction between online safety and data protection as they relate to age assurance.[154]

MRF also has strong concerns around the burden that relying on the age of digital consent places on parents. As it stands, many parents are confused about what the age of digital consent is, particularly in comparison to outright restrictions on children's access. Parents' ability to confidently engage with consent processes also varies significantly, particularly in the absence of supportive infrastructure liked paired social media accounts.

[153] Ofcom (2025) Children and Parents: Media Use and Attitudes Report
[154] Digital Regulation Cooperation Forum (2026) Age Assurance: A Joint Statement

Finally, in countries where a higher digital age of consent has been set, for example the Irish Republic and France, there is limited if any evidence that a higher digital age of consent has delivered meaningful improvements to children's access, safety or wellbeing.[155]

Time limits on children's access to online services

MRF is open to the introduction of time limits on children's access to online services, such as daily screen time limits or restricting overnight access for individual apps.

In particular, time limits are preferable to the introduction of an outright ban. Not only are they less likely to incur unintended consequences for vulnerable teens, they are less likely to have a chilling effect on safety-by-design and other on-platform regulatory outcomes.

Analysis from the University of Chicago also finds that time limits are more likely to achieve high levels of compliance and a shift in social norms, as they reduce the negative impacts of compliance (e.g. on children's social lives) and therefore the incentive to circumvent restrictions. They also more consistent with children's own wishes.[156]

In order for time limits to be effective, however, they must be implemented effectively. This is vital given significant outstanding questions around how these interventions would work in practice, with few examples of successful implementation in other contexts. As set out in Section 2, restrictions on overnight gaming introduced in South Korea ultimately did not lead to a reduction in time spent online or an improvement in key outcomes for children.[157] [158]

MRF has particular concerns about the likely effectiveness of any time-limit interventions introduced by services themselves. As seen elsewhere, existing safety tools designed by platforms are often ineffective, while early data from Australia demonstrates the consequences for compliance if platforms do not have to adhere to rigorous minimum design standards. As such, any time limits must meet rigorous and mandatory minimum design standards.

In this context, MRF strongly welcomes the 'IRL Trial' led by Professor Amy Orben and Dr Dan Lewer, which will provide valuable insights around both the implementation and potential outcomes of these interventions, and encourage the Government to make any decisions following the full results of this trial.[159]

Parental controls

MRF supports effective parental controls as part of a broad set of measures to deliver improvements in children's online safety and wellbeing. In order to ensure their effectiveness, they must be subject to a clear set of enforceable minimum standards, ideally operationalised via statutory codes of practice encompassing both platform and device-level tools.

---

[155] European Union Better Internet for Kids (2025) Protection children in the digital age: insights from France's education and regulatory initiatives

[156] Bursztyn, L (2026) Why Bans Fail: Tipping Points and Australia's social media man. University of Chicago, Becker Friedman Institute of Economics. Working paper No 2026-57

[157] Lee, C (2017) Ex post valuation of illegalising juvenile online gaming after midnight: a case of shutdown policy in South Korea. Telematics and Informatics, 34(8)

[158] Choi, J et al (2018) Effect of the online game shutdown policy on Internet use, Internet addiction, and sleeping hours in Korean adolescents. Journal of Adolescent Health 62(5)

[159] University of Cambridge (2026) Thousands of UK schoolchildren to take part in major study of social media use and teen mental health.

If designed and deployed effectively, parental controls can offer clear benefits to children's safety and wellbeing. We note that parents show strong support for greater oversight of their children's online behaviour. When asked their preference between 'stronger parental control tools built into devices and apps' and a social media ban, 55% of parents supported parental controls, compared to 39% who supported a ban.[160]

However, we cannot rely on voluntary action by platforms, even if underpinned by new guidance. As it stands, MRF has serious concerns about existing parental controls, with take up low, effectiveness patchy, and evidence suggesting that overly intrusive controls may worsen rather than improve safety outcomes. MRF research on the 47 safety tools introduced by Instagram's Teen Accounts identified that almost two thirds (64%) were either unavailable or ineffective (able to be trivially circumvented or evaded during red-team testing). A further proportion (19%) offered some functionality but came with notable limitations and only 17% working as advertised. There were specific issues with parental controls. Despite Meta press-releases to the contrary, safety testing demonstrated that parental supervision features were user-activated rather than default; did not accurately present what a teen was seeing on their account, including when they had been recommended harmful suicide and self-harm content; failed to notify parents if their child reported a post or account; and could easily be entirely circumvented by a child setting up a secondary account on their device.[161]

Many parental controls are also extremely hard to use, being difficult to find, confusing, or hard to set up. This is a clear example of 'friction by design', with services deliberately seeking to frustrate users from turning on features and settings that may reduce children's engagement and therefore not be in the platform's commercial interest.[162] [163]

These failures are symptomatic of a wider pattern of services using the introduction of new safety features as a performative, headline-grabbing gesture to reassure parents that they have their children's interests at heart, while clearly failing to deliver meaningful improvements in safety and wellbeing outcomes.

In this context, we strongly urge the Government to apply robust minimum design standards to ensure that parental controls are appropriately targeted, on by default wherever appropriate, easy to use, standardised across platforms and devices, and supportive of key safety and wellbeing outcomes.

We would also caution Government that, due to variation in parents' capacity to effectively use controls, they must only be part of a wider package of interventions to support children's safety and wellbeing. As set out in the consultation, parents' digital and media literacy varies significantly across different groups and domains, with many parents finding it difficult to keep up with their child's behaviour and offer appropriate support, having limited awareness of existing tools, and lacking confidence to use them effectively. Internet Matters data, for example, finds that only around a third of parents use tools like screentime management, gaming parental controls, and safe search settings.[164] There is also variability in the reasons

[160] Tech at Public First (2026) To Ban or not to Ban
[161] Molly Rose Foundation, Arturo Bejar and partners (2025) Teen Accounts, Broken Promises: How Instagram is failing to protect minors
[162] Ibid.
[163] Behavioural Insights Team (2025) Behavioural Audit of Online Services.
[164] Internet Matters (2025) internet Matters Pulse

parents will set up a control, with only 34% of parents taking action to set up a filter or parental control in response to their child encountering upsetting content. [165]

Virtual Private Networks (VPNs)

The Government is right to raise the risks of children and young people circumventing age limits through the use of Virtual Private Networks (VPNs).

As it stands, it is important to note that there is no strong evidence that large numbers of young people are using VPNs to get around age restrictions, either in the UK or Australia.

MRF polling identified that only around 5% of Australian 12-15 year olds who had retained access to accounts on restricted platforms had done so by using a VPN, with the vast majority not needing to find workarounds as their existing accounts had not been removed.[166] In the UK, research found 'no spike' in VPN usage following the OSA's age verification requirements coming into force in July 2025.[167]

Looking ahead, however, Government should look to restrict children's access to VPNs, bringing app stores into scope of the Online Safety Act and applying age checks if children wish to access them.

[165] Ofcom (2025) Children and Parents: Media Use and Attitudes Report
[166] Molly Rose Foundation (2026) Australia's social media ban – is it working?
[167] Childnet (2025) Young people's use of VPNs

Section 7: Further action on online safety

Summary

MRF welcomes the Government's ongoing commitment to ensuring children live safe and healthy online lives, and the proposals currently being consulted on are a welcome signal of intent.

However, any new measures must be a downpayment on further action to deliver the comprehensive reset that children and families demand and deserve.

It is clearer than ever that preventable online harm is first and foremost an issue of product safety. As in every other part of the economy, the answer to market failure is strong and effective regulation.

The Online Safety Act is therefore a crucial part of the solution. However, after years of delays and watering down, its implementation has fallen badly short. Ofcom oversees a triangulated regime that fails to tackle the incentives and business models that perpetuate harm and continue to cost young lives.

A strong and ambitious regulatory fix will ultimately be required – one that can comprehensively address the appalling preventable harm faced by children online, shift the underlying incentives and business models that continue to treat children's safety and wellbeing as an afterthought, and deliver a regulatory regime that is commensurate in scope the largest and most cash-rich companies in the world.

This can be achieved through a combination of immediate and systemic action:

- In the current parliamentary session: a swift and decisive legislative package that can immediately address the most pressing structural barriers to robust and active regulatory enforcement
- In the third parliamentary session:
    - Comprehensive legislation to fix and strengthen Act's foundations, including introducing an overarching Duty of Care, resetting regulatory incentives in favour of harm reduction, and delivering an outcomes- and conduct-based regulatory approach
    - Extending the scope of the Act to cover wellbeing, making wellbeing-by-design the price of entry to the UK market

Both the current and previous Secretary of State have acknowledged that delivering for children and families will ultimately require further legislation.[168] [169] This also reflects the preferences of

[168] Adu, A (2025) UK online safety laws 'unsatisfactory' and 'uneven', says science minister. Published in the Guardian, 12th January 2025

[169] Paul, J. (2026) Liz Kendall insists Government will 'legislate every year' to keep up with the pace of technology. Published on LBC, 9th March 2026

parents, with three quarters (73%) of UK adults support new legislation to strengthen regulation and better protect children and young people from harm,[170] with support for regulation stronger than recent support for an Australia-style social media ban.[171]

Immediate measures to fix the Act

In the current parliamentary session, the Government should commit to a series of immediate measures, through a set of targeted technical amendments, to address some of the major structural issues within the Act — issues which Ofcom has said are hindering its ability to deliver the regulatory outcomes Parliament originally intended.

If the Government acts quickly to address the barriers that Ofcom says it is facing – while it also develops and legislates for a bolder and longer-term set of changes – this will send a powerful message to both parents and the public that change is on the way,

This should include:

- **Removing the 'clear and detailed' and 'technically feasible'** requirements that have throttled the ambitious of Ofcom's Codes of Practice – preventing the stretching and outcome-focused set of measures that Parliament had anticipated, and meaning current codes do little more than bake in the status quo.
- **Removing the 'safe harbour' principle** that allows some platforms to legitimately claim regulatory compliance while still being able to scale back their existing safety efforts. In combination with the limited ambition of the codes, this is having a chilling effect on safety by design innovation, and means regulation is acting as a ceiling, not a floor, for children's safety.
- **Mandating safety-by-design,** by inserting a clear definition in the Act and requiring the introduction of a safety-by-design code of practice. MRF supports the model code which has been prepared by the Online Safety Act Network alongside MRF and other leading organisations.[172] This would ensure that platforms have to address the harmful design choices, features and functionalities that drive harm, and introduce provisions for regulated firms to rigorously test their products and to ensure they are demonstrably safe for use before being released to the public.
- **Strengthening the obligation in the Act that platforms take reasonable steps to reduce the risk of harm to users, as identified in their risk assessments.**

A full suite of amendments has been prepared by the Online Safety Act Network, which MRF wholeheartedly endorses.[173]

A new Act that substantially strengthens the regime, refocusing it on harm reduction, outcomes and conduct

Looking ahead, the Government must then make a commitment to a White Paper and announce that a new Act is on the way in the third parliamentary session.

---

[170] Research conducted by Savanta for Molly Rose Foundation. 2,048 UK adults were surveyed in January 2026.
[171] Recent polling for The Mirror found support for a ban at 66%
[172] Online Safety Act Network (2026) Safety by design code of practice: parliamentary briefing note
[173] Online Safety Act Network (2026) A 10-point plan for Government

A new Act must demonstrate a meaningful shift in regulatory approach, with a new set of measures that can decisively tackle the underlying incentives and business models that continue to fuel the entirely preventable harm faced by teens. This must include:

- **Introducing an overarching Duty of Care**. This would fundamentally shift the onus onto tech firms – setting a clear requirement for them to address and respond to reasonably foreseeable harms, and incentivising a truly systemic approach to risk identification and mitigation. This also moves us away from the regulator needing to play whack-a-mole in the face of widespread non-compliance, and matches comparable approaches in other sectors.[174]
- **Introducing a new conduct-based approach** that is commensurate to the companies and problems in scope of the Act. Learning from the financial services sector, regulated companies and senior responsible staff should be required to conduct themselves with due regard to the regulatory system, with new measures also requiring a culture that promotes and upholds child safety and wellbeing, transparency, and the best interests of the child. This should be actively underpinned by a robust senior manager scheme, resetting the incentives at both entity and individual level.
- **Focusing the regime on measurable harm reduction**. Section 1 of the Act should state unambiguously that Ofcom's primary objective is to reduce exposure to online harm. There should be a parallel statutory obligation for Ofcom to prioritise children's safety over industry growth. The regulator should also face a new overarching duty to deliver measurable reductions in exposure to harm among young people, and to write to the Secretary of State if these targets are missed.
- **Resetting the regime in favour of victims:** The regulator should face an explicit legal duty to protect the fundamental right to life (Article 2) and to demonstrate how it has upheld its positive obligations when deciding on and implementing safety measures and policies.

Extending the Act to cover Wellbeing-by-Design

A new Act should also expand the scope and ambition of the Act – widening its objectives to not only necessitate harm reduction but to actively deliver wellbeing-by-design.

As explored in this consultation, this would acknowledge and respond directly to parents' profound concern around the chronic risks of time spent online – enabling regulation to tackle the harmful, persuasive and compulsive design features that incentivise problematic and excess product use.

A fundamental reset of the expectations we place on tech firms is required. Through a combination of new duties on platforms, and new strategic objectives on Ofcom, Government should introduce a robust and demanding set of measures that make clear that children's wellbeing is the price of admission to the UK market. Extending the Act in this way would mark an end to addictive and harmful design choices, and ensure digital products are built to be age-appropriate, high quality and nourishing by design.

An expanded Act should mirror the strengthened approach elsewhere in the regime, and should include:

[174] Most notably the introduction of the FCA's Consumer Duty

- **A new overarching duty on platforms to ensure they are designed to protect and promote the wellbeing of children and young people.** This duty should apply not just to social media, but to gaming companies, messaging platforms and AI products. Platforms should also be required to conduct a wellbeing-by-design impact assessment, doing so to a suitable and sufficient standard.
- **New statutory objectives on the regulator to deliver wellbeing-by-design when discharging its functions.** The regulator should be prepared to measure changes in wellbeing over time, using evidence-based approaches to appropriately capture a range of primary wellbeing measures.
- **A new statutory Code of Practice on wellbeing and addictive design**. This new code should ensure that harmful and addictive design features are prohibited, and must confront the business models that drive chronic harms around compulsive behaviour and the opportunity costs of time spent online.

As explored in this response, an expanded Act and statutory Code of Practice on wellbeing and addictive design should include robust requirements around requiring algorithmic feeds to promote trusted, high-quality content and meet content plurality requirements.

It should also prioritise a range of broader positive wellbeing-by-design objectives, drawing on thinking from child rights, academic research, and evidence for the positive impacts of online spaces. Priorities should include supporting meaningful agency over online experiences, and robust requirements to promote positive and healthy social connection.

Appendix 1: Open Space event and feedback – How do we keep young people safe online?

Context

On March 19th Molly Rose Foundation, alongside Flippgen and Beyond, held a unique event that brought together Government, Ofcom, civil society, young people and those with lived experience of online harm to discuss how we keep young people safe online. The event, known as an 'Open Space', allowed delegates to set their own agenda and discuss any topic they felt important to the question. It put young people and those with lived experience on an equal footing with regulators and policy makers.

Attendees included year 6 pupils as well as sixth formers and young adults, alongside representatives from DSIT, Home Office and DfE and Online Safety Minister Kanishka Narayan. Representatives from civil society included the Conscious Advertising Network, NSPCC and many others, including grassroots organisations and larger online safety charities.

The event saw three sessions with multiple discussions around the question: 'How do we keep young people safe online?' It was held under Chatham House rules so the feedback below has been anonymised.

We encourage Government to consider the key themes raised by attendees alongside MRF's wider consultation response.

Key themes

The event was held in a way that anyone could propose a topic to discuss and host it, so it was entirely democratic. Many of the discussions and questions were posed by young people and those with lived experience. There was a diversity of topics but some key themes emerged, including the impact of social media, education, AI, social media bans, the role of advertisers, age verification, responsibility and bullying.

We have summarised some of the conversations below, but this is just a snapshot of the day and key learnings were taken away by Government departments, Ofcom and the Minister himself who engaged with young people in a number of discussions.

Impacts of time spent online

Discussion were held on the impact of social media and other online platforms, with participants referencing both positive and negative aspects of time spent online.

In a group made up almost entirely of sixth formers discussing perfection and social media, the group reflected that social media could introduce feelings of pressure to be perfect. Some said social media made them feel like they aren't normal or look at themselves negatively. However, others said social media was a way for young people to express themselves. The group agreed that it's difficult to find a balanced middle ground on social media. They also reflected that

reality can be blurred on social media with difficulty understanding what is real and fake when online.

Discussing bullying, one group reflected on how this no longer ends in schools and can come in the form of direct messaging and on public posts or in group chats. They said that hidden identities can fuel online bullying and how social media comments can linger for life. However, they did reflect that online bullying is not necessarily more impactful or important than bullying face to face.

Age verification

When discussing age verification, it was clear that effective age verification is needed but discussion ranged on what that looks like. There was a consensus that age verification should exist but it needs to be more effective than the current measures. One 11-year-old reported that Roblox verified them as 18 despite their young age. There were also concerns around data security and how age verification works across different jurisdictions (e.g. if a server was in another country). They also questioned what age verification should be applied to, including whether it should it be at app store level and whether it should apply to individual features.

The group also debated the role of schools in educating young people about why age verification exists and how their data would be used. One young person reflected that 'transparency is so important'. In general, the group felt there was a gap in the education offered to both young people and parents on the implementation of the Online Safety Act. The young people said they did not notice the other impacts of the Online Safety Act (e.g. on harmful content), so the only visible effect was requiring age verification

Social media bans

On the topic of social media bans, the groups raised whether such a step might actually incentivise young people to use restricted platforms that might be less safe, and queried what children would replace time spent online with if there was a well-enforced ban, citing a lack of offline activities for young people. The idea of 'forbidden fruit' was raised and they cited a lack of evidence as to whether bans work. In a straw poll the consensus was that a ban was not the answer.

Instead risk-based age ratings were discussed and how this might incentivise 'an arms race towards better'. They again cited the importance of digital literacy and critical thinking across the board.

Education

There was widespread discussion about online safety education, with a consensus that online safety education needs a reset to reflect children's real digital experiences. Current approaches are fragmented and increasingly challenged by parents, despite clear demand for earlier and more effective provision. Strong practice exists in whole school, practical and peer led models, but key gaps remain in areas like algorithmic literacy, open discussion, and alignment with digital parenting. Delegates explored the need to move beyond fear-based messaging towards a coherent, connected approach that builds skills early, uses everyday technologies as entry points, and equips children and families to navigate online environments confidently.

Tensions included online safety being taught in siloes, largely due to a siloed national curriculum. Parents also seemed to want early education but there were concerns about their

children being introduced to social media at primary age even if this was done via structured lessons rather than solo use. There were suggestions that online safety is becoming the new 'sex ed' with parents looking to ban and remove their children from lessons. This includes a backlash against ed tech. Lessons could be learnt from where schools have become braver with previously taboo topics like RSHE.

In terms of what's working, the group cited project-based learning and peer education with young digital leaders. They also said daily messages in a whole school environment could reinforce safety, while practical do's and don'ts would be useful as well as high quality resources that span the curriculum.

Key gaps included a lacked of training around how to have agency over algorithmic feeds, a lack of space to discuss online experiences, and a lack of guidance for parents on how to channel fear into meaningful action. It was agreed that there is a need to stop demonising technology to children and remove blame, shame and moral panic. Instead delegates prioritised filling this space with credible and diverse voices, as well as 'connecting the curriculum' (for example so that algorithmic literacy would be taught in computing).

Responsibility for online safety

Discussions were held about whose responsibility online safety should be, with a separate session also held on whether parents should take more responsibility for children's safety. The general consensus was that online safety is everyone's responsibility - including Big Tech, parents and carers, Government, schools and regulators.

Delegates reflected that Big Tech should focus on age gating and content moderation overseen by humans as well as restricting addictive design features. Government should also enforce policy with teeth and ensure it has an appropriate distance from the influence of Big Tech.

While parents and educators hold responsibility, they need support. There was a discussion about how education and resources could be offered to both parents and young people. Some delegates suggested a more 'joined up' approach between schools and parents to ensure a consistent approach to safety.

Other issues raised

Some young people reflected that they were not aware of Ofcom or the work they do, and felt unrepresented by the regulator. They called for Ofcom to do more by engaging with young people on social media and setting up a youth board to help get young people's perspectives on regulation. They also called for more transparency from Ofcom to tell the story of regulation and what changes are being made to protect children.

Young people also raised interesting questions about parental controls, suggesting that these inhibit their natural curiosity, and instead suggested that digital literacy was a better solution and less likely to send them in potentially dangerous directions. To this they also called for more education in algorithmic literacy, as well as critical thinking across the board.

A session was held on managing the Manosphere with a consensus that education needs to be consistent and start in primary school. This should include education on algorithms. Delegates also called for industry action to tackle how algorithms drive and monetise misogynistic content. There were also calls for practical action Government and regulators could take, such as making the VAWG online safety code compulsory for tech firms.

There was also concern raised about AI, in particular because of perceived lack of effective regulation. There was a sense that strengthening education around AI would be hugely beneficial.

A session was held on the role of advertisers in online safety and it was agreed that more transparency is needed across the industry, with adverts monetising harmful content. It was agreed more transparency is needed, for companies themselves as well as regulators and Government. The work of Conscious Advertising Network was flagged and the need to incentivise companies not to have adverts placed next to harmful content.

Sessions were also held on feminism, intersectionality and the role of online gaming. Discussions on online gaming referenced that this a key space for young people to connect and have fun, and any new safety measures must include gaming platforms in their scope.

Conclusion

This document is a short snapshot of the discussions held at the Open Space event. Overall, delegates called for decisive action by Government to better protect young people online.

The consensus was clear that too long has passed without meaningful action on online safety. It was felt that the consultation is a pivotal moment to get this right, and by listening to the voices of young people and those with lived experience the Government has an opportunity to do just that. If the Government acts decisively and in an evidence-led manner, it will have the backing of civil society, young people and those with lived experience.

It was also noted - as mentioned above - that Ofcom and other decision makers can do much more to engage diverse voices. Molly Rose Foundation, founded in lived experience, is ready to help facilitate this and ensure that the voices of young people and those with lived experience help drive the urgent change that's needed.

Appendix 2: Supporting information

Are you answering as a private individual or on behalf of an organisation?	On behalf of an organisation
Organisation name	Molly Rose Foundation
Organisation type	Civil society / third sector / community
Location	In the UK
As part of your current occupation, do you work with children aged 21 or younger in an education setting?	No