

Systematic failures in Online Safety Act risk assessments – why is Ofcom failing to act?

Policy briefing, December 2025

Risk assessments are a cornerstone of the Online Safety Act, with online services required to produce ‘suitable and sufficient’ risk assessments for both the illegal and child safety parts of the regime. However, new Ofcom analysis suggests there were ‘notable issues’ with many of the first assessments produced.

MRF is concerned that online platforms are systematically downplaying the risks on their products in their risk assessments. Remarkably, not a single social media platform has determined they are high-risk for suicide, self-harm and depression content.

Ofcom has failed to respond effectively. Despite widespread issues, the regulator has declined to act against a systemic trend of online platforms systematically downplaying the harm they cause to children. **Ofcom’s approach risks letting platforms off the hook – and sends a clear signal to tech firms that the regulator is reluctant to act against even flagrant breaches of regulatory duties.**

Our concerns

1. Risk assessments are failing to accurately capture harm levels – and Ofcom is failing to adequately act

Risk assessments are designed to be a crucial part of a systemic regime – with platforms expected to assess for risks and determine whether they are high-, medium- or low-risk for the Priority Offences and the range of children’s harms in scope. However, Ofcom’s early analysis suggests that many platforms have palpably failed to undertake this process rigorously.

The regulator has identified a litany of issues with initial risk assessments¹ - many platforms may have inaccurately assessed illegal and harmful content, poorly accounted for risks resulting from their design choices, and two-thirds failed to name a person responsible for ensuring the risk assessments are robust. **One major platform even claimed that it was low-risk for every single harm it analysed.**

¹ Ofcom (2025) Year One online safety risk assessment: report on standards and improvements. Published December 4, 2025 <https://www.ofcom.org.uk/siteassets/resources/documents/online-safety/research-statistics-and-data/os-standards/online-safety-risk-assessments-report-year-one.pdf?v=408834>

Despite the regulator setting out that it still has ‘outstanding concerns over the suitability and sufficiency of several providers illegal and children’s risk assessments’ – in other words, that it has observed clear-cut breaches of the Act – Ofcom is still to take enforcement action.

In respect of the site above, Ofcom states it continues to undertake ‘compliance remediation discussions’ with them.

2. Platforms are systematically downplaying their suicide and self-harm risks

Astonishingly, not a single platform claims that it is high risk for suicide, self-harm, depression eating disorder content, with the vast majority (out of 34 risk assessments disclosed) claiming they are low risk for suicide and self-harm material. The regulator said several providers with recommended systems failed to explain how they mitigate the risks of cumulative exposure to harmful content, including the suicide and self-harm material that cost Molly’s life.²

These risk ratings are simply not credible - and it is hugely concerning Ofcom has not done more to challenge them.

At the same time that children’s risk assessments were being produced, MRF undertook research which found that over 95% of algorithmically recommended posts on a 15-year old’s Instagram and TikTok account contained harmful suicide, self-harm and intense depression content.³

Separate research found that more a third of teens aged 13-17 – and almost half of girls – had been exposed to suicide, self-harm, depression and eating disorder content across six of the largest sites in the previous week.⁴

3. Platforms are downplaying their suicide offence risks

Only one of 69 Illegal Risk Assessments claim that their platform has a high risk of suicide offences – that is, encouraging or assisting a suicide act. This flies in the face of the rapid growth of grooming for purposes of suicide and self-harm by organised groups, known as Com groups, with online offenders identifying and grooming vulnerable teenage girls on social media platforms, gaming platforms and forums such as Discord.

The FBI has warned about the ‘sharp increase’ in the threat posed by these groups, with similar warnings from the National Crime Agency, Europol, Canadian and Australian law enforcement.⁵

This is the latest evidence that Ofcom seems unable or unwilling to grasp the threat posed by suicide offences (MRF has previously been highly critical of Ofcom’s failure to adequately account for the threat posed by Com groups in either its Register of Risks or codes).⁶

Ofcom states that it asked 11 platforms to revisit their risk assessment, and five platforms to reconsider their

² Ofcom has signalled that this will be a priority in 2026, with a further update in May.

³ Molly Rose Foundation (2025) Children’s Exposure to Suicide, Self-Harm, Depression and Eating Disorder Content Online. <https://mollyrosefoundation.org/wp-content/uploads/2025/10/Molly-Rose-Foundation-childrens-exposure-to-harmful-content-1.pdf>

⁴ Molly Rose Foundation (2025) Pervasive-by-design: suicide, self-harm and depression content on TikTok and Instagram, and how their algorithms recommended to teens https://mollyrosefoundation.org/wp-content/uploads/2025/08/proof3_PervasivebyDesign.pdf

⁵ FBI (2025) Violent online networks target vulnerable and underage populations across the United States and around the globe - Public Service Announcement. Published March 6th, 2025

⁶ For more background, see an op-ed written by MRF in March 2025: https://www.lbc.co.uk/article/com-groups-suicide-and-self-harm-online-radical-gamer-ofcom-5Hjd2qN_2/

risk ratings. As a result, it **appears that Ofcom has tacitly endorsed the risk ratings adopted by the majority of platforms to date – a position that seems wholly divorced from the evidence that was available at the time.**

Why this matters

Risk assessments are a crucial part of the regulatory regime, with suitable and sufficient risk assessments an integral component of the systemic approach to identifying and mitigating risks intended by the Act.

In the face of clear-cut and seemingly flagrant breaches of the Act, Ofcom's decision to primarily rely on supervisory discussions, rather than swift and publicly visible early enforcement action, sends a worrying message to tech companies about the regulator's willingness to move swiftly to enforce systematic – and even potentially coordinated - regulatory breaches.

Ofcom's approach compounds previous decisions that could reasonably have been interpreted as evidence of the regulator adopting a 'light touch' approach to its enforcement strategy. Earlier this year, Ofcom stated it was relaxed about Meta using generative AI to produce its risk assessments, so long as the 'suitable and sufficient' standard was met. (At the time, MRF and 23 others have warned of the risks that platforms could use to game their risk assessment duties.)⁷

MRF continues to be desperately disappointed by the lack of ambition in the regulator's approach. **Last week, the regulator quietly announced its intention not to bring forward much-needed further measures to strengthen its codes of practice until at least late 2027 - meaning any further duties would not be enforced until late 2028 at the earliest.**⁸

⁷ Letter to Ofcom CEO Melanie Dawes, sent June 6th 2025 <https://mollyrosefoundation.org/wp-content/uploads/2025/06/risk-assessment-letter-June-25.pdf>

⁸ Ofcom (2025) Online Safety in 2025: a summary of the technology sector's response to the UK's new online safety rules. Published December 4th 2025 <https://www.ofcom.org.uk/about-ofcom/our-research/online-safety-in-2025-summary-of-the-technology-sectors-response-to-online-safety>