

Dame Melanie Dawes
Chief Executive

22 October 2024

Dear Secretary of State,

Thank you for your letter of 16 October. I share your commitment to ensuring that the Online Safety Act (the Act) is implemented as quickly as possible. As you know, we published an update on our progress on 17 October, which confirms that we are on track to meet the timetable that Parliament set for us. The end of this year marks a turning point. In December, we will publish the final rules on illegal harms and will be turning the spotlight on to services as their duties start to kick in. All firms in scope of the new rules must act, and we will expect them to do so swiftly to deliver a safer life online for people in the UK.

The hateful and violent disorder seen in the aftermath of the horrific attack in Southport this summer is a matter of great concern to us both. The Act places new legal duties on platforms available in the UK, including social media sites and file sharing apps, gaming and messaging services, and search engines. They will be required to take steps to prevent the spread of illegal material and mitigate the risks to the safety of people in the UK arising from illegal online activity, and content and conduct which harms children. This includes activity that provokes violence or stirs up hatred, and false communications intended to cause harm.

From the end of this year, firms must initially assess the risks associated with illegality on their platforms, and then, once our Codes of Practice are in force, use the measures we have recommended, or other effective measures, to mitigate these risks. Our Codes and guidance will require that firms address safety risks in the design of their products, ensure their terms of service prohibit activity that is illegal under UK law, and remove illegal material once they become aware of it.

As you requested, I am writing to set out our evaluation of how illegal material may have spread during the period following the Southport attack, the actions online services took in response, and how we will expect firms to respond to similar events once the new duties are in force.

Immediate steps in August

Ofcom acted swiftly after the violent disorder began, issuing a public statement on 5 August, followed by an open letter to tech firms, emphasising the importance of taking steps to protect users and reminding them that they did not need to wait for their new duties to come into force to do so.

In parallel, our Supervision teams spoke directly and in detail with many of the largest social media and messaging platforms from early August, to find out their response to the riots. We also met with a range of stakeholders from civil society, Government and law enforcement. It is clear that critical incidents such as this will require close working with Government and other public bodies, and we

are grateful for the rapid input we received from a wide range of independent and expert organisations.

To ensure we had a clear understanding of events, information was provided to us confidentially, and as a result we will not share details of individual companies' responses. And given the duties in the Act are not yet in force, we have not sought to determine whether they would have been breached by what we saw. Once our powers are in force early next year, we will be able to consider whether to open an investigation into any future incidents of this kind, and take enforcement action if we have evidence that a breach of the rules may have occurred.

Nonetheless, these events have clearly highlighted questions tech firms will need to address as the duties come into force. While some told us they took action to limit the spread of illegal content, we have seen evidence that it nonetheless proliferated, and appears to have contributed to the significant violent disorder which followed the tragic murders in Southport. Of the numerous convictions which have followed, some have been found guilty of online posts threatening death or serious harm, stirring up racial hatred, or sending false information with intent to cause harm.

The role of online platforms following the attack

Our assessment is based on information provided to us by tech firms and other stakeholders. Our key conclusions are:

1. Illegal content and disinformation spread widely and quickly online following the attack

While not all platforms experienced significant levels of illegal and harmful material, others have confirmed to us they were dealing with high volumes, reaching the tens of thousands of posts in some cases. Misinformation appeared online almost immediately after the attacks, some of it appearing to have malicious intent and seeking to influence public opinion and reaction. We have heard how some platforms were used to spread hatred, provoke violence targeting racial and religious groups, and encourage others to attack and set fire to mosques and asylum accommodation. Accounts (including some with over 100,000 followers) falsely stated that the attacker was a Muslim asylum seeker and shared unverified claims about his political and ideological views. Posts about the Southport incident and subsequent events from high-profile accounts reached millions of users, demonstrating the role that virality and algorithmic recommendations can play in driving divisive narratives in a crisis period.

2. There was a clear connection between online activity and violent disorder seen on UK streets

Some major platforms were used to post and disseminate material calling for violent action in Southport and in other towns and cities. One analyst told us that calls for demonstrations targeting a local mosque were circulating in private groups online within two hours of the vigil for the victims of the attack. The evidence we have seen indicates to us that some messaging services hosted closed groups comprising thousands of users in some cases. Some of these groups disseminated material encouraging racial and religious hatred, and provoking violence and damage to people and property, including by identifying potential targets for damage or arson.

3. Most online services took rapid action in response to the situation, but responses were uneven

Firms told us they took a range of actions in response to these events, including:

- The implementation of incident response protocols by some services. This included setting up internal cross-functional groups focused on the crisis, monitoring spikes in relevant types of harmful content, and the blocking of URLs leading to illegal and harmful content on other sites.

- Actions to prioritise the moderation and removal of illegal and harmful material linked to the protests, and to suspend or close down accounts and channels involved in spreading such content.
- Proactive engagement with civil society and/or law enforcement partners to seek guidance. In one instance, a service proactively reached out to a civil society organisation focused on anti-Muslim hatred requesting training to improve their moderation systems and took down potentially illegal content based on a referral from law enforcement agencies.

Questions for platforms

Our proposed Illegal Harms Codes and Risk Assessment Guidance set clear standards for what we will expect to see in future. These events highlight questions that a number of platforms would have had to answer, had the duties been in force at the time. Specifically, the draft Codes and Guidance propose that user-to-user services should:

- a) Include in their **terms of service** provisions specifying how individuals are to be protected from priority illegal content. In any future incident, we will be concerned if services do not have **clear and consistent** terms of service prohibiting illegal hateful content. It is not clear that this is currently the case on all services.
- b) Have systems or processes designed to **swiftly take down illegal content** once aware of it, by **prioritising the moderation of viral content**, and having content moderation teams that are **adequately resourced and trained to meet increases in demand caused by crises**. On some platforms, false information regarding the identity of the attacker continued to spread in the three days it took for his real identity to be made public, even when there was evidence of intent to stir up racial and religious hatred. Once the duties are in force, tech firms will need to act promptly when they become aware they are hosting such material, and we will want to understand their targets for removing it, and whether they are meeting them. Moderation of livestreamed material is particularly important and challenging, given its ephemeral nature, and the use of this functionality during this period to spread hateful messages and provoke violence.
- c) Provide effective and accessible mechanisms for **users to complain about illegal content**, including material that stirred up hatred, provoked violence or constituted violent threats. Given the volume of potentially illegal content which appeared during this period, we will want reassurance that platforms' reporting systems are easy to find and use, and that they can cope with the number of user reports they receive, especially in crises, and moderate flagged content swiftly. This includes **messaging services**, especially those hosting groups comprised of large numbers of individuals.
- d) Assess whether **changes made to recommender systems** are likely to increase users' exposure to illegal material. Given evidence that illegal and harmful content linked to these events was recommended and amplified by some services, we will expect services to be able to show how they test their recommender systems following any changes to ensure they don't expose users to additional risk.

In addition to these protections from illegal material for all users, our draft Protection of Children Codes require that sites and apps likely to be accessed by children take additional steps to ensure that children are **protected from material which, while not illegal, is harmful. This includes incitement to violence and hatred as well as hateful abuse**. We have made clear in our draft

proposals that algorithms need to be changed so that this type of content is **downranked on children’s recommender feeds**.

On this basis, I am confident that, had the draft Codes been in force at the time, they would have provided a firm basis for urgent engagement with services on the steps they were taking to protect UK users from harm.

Next steps

Ofcom will use these findings to drive our engagement with regulated services ahead of the duties coming into force in next year.

We will also use this work to examine whether there are further measures we should add to the future iterations of Codes. Specifically, as part of the consultation on additional measures for the Codes in Spring 2025, which we announced in our update to Parliament last week, we will bring forward a new measure that would place additional requirements on services regarding their crisis response protocols. We are also considering the evidence base for potential new measures on livestreaming functionalities and recommender systems.

These events also highlight the importance of promoting media literacy, to heighten public awareness and understanding of how people can protect themselves and others online. We recently set out a new media literacy strategy,¹ including explaining how we will work with stakeholders to support the users of regulated services to understand the nature and impact of disinformation and misinformation, and to reduce their and others’ exposure to it. Google Search, The LEGO Group, Pinterest and Roblox have already voluntarily adopted Ofcom’s Best Practice Principles for Media Literacy by Design and we have invited other online service providers to publicly demonstrate their commitments to media literacy by doing the same.

We will shortly be establishing the new Advisory Committee on Misinformation and Disinformation, and this will provide a further source of expert advice on how tech firms should deal with this type of harm, how Ofcom’s transparency powers could be used, and on the role of our media literacy duties. We anticipate advertising for Committee members before the end of 2024, so the Committee can start work early next year.

I look forward to updating you and Parliament as this work progresses.

Yours sincerely,

Melanie Dawes

¹ [A Positive Vision for Media Literacy: Ofcom’s Three-Year Media Literacy Strategy](#), October 2024