

## Your response

Question	Your response
<p><b>Question 1:</b> We include labels, overlays, pop-ups, notifications, and resources as examples of on-platform interventions (additional information regarding this typology can be found in the Annex on page 3).</p> <p>(a) Do you agree with this categorisation of on-platform interventions?</p> <p>(b) If not, please explain.</p>	<p>Confidential? – N</p> <p>(a) The Trust agrees with the categorisation of interventions.</p>
<p><b>Question 2:</b> Do you have any feedback on the summary of themes we identified from online services? Are there any omissions or other items you think important to add?</p>	<p>Confidential? – N</p> <p>We agree that the themes identified are important and that interventions should be guided by impact assessments and evidence regarding which interventions works best, in what context and timing, to create a safer environment for users of social media platforms.</p> <p>Exposure to disinformation, racism and other harmful content which is routinely found online will make it harder to combat it when a major event happens. We therefore believe that interventions should be routine and not only when external events such as wars or pandemics amplify harmful content.</p>
<p><b>Question 3:</b> Are we missing anything with the three headings used to structure the best practice principles for media literacy by design?</p>	<p>Confidential? – N</p> <p>The headings are comprehensive. We would add that platforms should be proactive about their interventions and evolve with new and emerging threats.</p>
<p><b>Question 4:</b> Which aspects of the proposed best practice principles for media literacy by design work well, and why? Which aspects don't work so well, and why? Do you have any comments on the specific principles (please specify if providing feedback on individual principles)?</p>	<p>Confidential? – N</p>

**Question 5:** Do you have any further guidance/feedback to offer on how platforms can enact best practice media literacy by design?

Confidential? – N

The Trust also urges platforms to have interventions in place for content generated by Artificial Intelligence. The quality of AI-generated content is advancing rapidly, and the threat posed by high-quality, mass-produced, and widely disseminated misinformation is a matter of significant concern. AI enables anti Jewish incitement through the creation of sophisticated antisemitic fake images and ‘deepfake’ videos.

The Community Security Trust (CST) found for example that users on extremist platform 4Chan have produced antisemitic images using generative AI tools. Users asked AI to produce an image of ‘Jew about to be killed, afraid, screaming.’ The image produced also included antisemitic stereotypes to make a character appear ‘Jewish’.

There should be labels in place to make user aware that an image or a video is not real, but AI-generated to help minimise the spread of disinformation. If AI-generated content shows harmful material, such as racist stereotypes this should also include an intervention, the same approach as for images that are not generated by AI.

**Question 6:** Can you submit any case studies or examples of different services enacting any of these best principles for media literacy by design? Can you provide any other examples of best practice media literacy by design that may not be covered by this document?

Confidential? – N

The Trust has worked with TikTok and X, to take but two examples, on educational initiatives. We worked with a company named Blue Robot to ensure that users interacting with a particular post had follow-up messages directed to them, and we worked with TikTok to produce content in relation to Holocaust Memorial Day. The company had both a headline (welcome) video, and a page with a collection of curated videos for the day. It took other steps in relation to searches for content about HMD. Behind this work are principles of sourcing, promoting and highlighting authoritative information over disinformation and hate. The issue is that these projects have been in isolation and need consistency of approach, investment and review.

**Question 7:** How do you expect in-scope services to demonstrate that they have adopted the principles? What would this look like?

Confidential? – N

So far platforms have rarely disclosed how effective their interventions are (and some platforms may not know it themselves). It's crucial that services should be transparent about their interventions and how effective they are by making the information publicly available and easy to find. The types on interventions and their aims could be permanently available on the platform. The effectiveness of these interventions could be submitted to Ofcom periodically, including plans to improve effectiveness.

**Question 8:** What more can be done to encourage services to promote media literacy by design?

Confidential? – N

Ofcom could publicly name companies who have proven to have good practices and have made efforts to keep users safe and create a positive and inclusive environment on their platform.

We also think that interventions on specific harms will not be enough to keep users safe. In addition to those measures platforms should employ other safety by design features that will make their products safer to begin with. This will help interventions be more effective, and it may also reduce the number of interventions required. This should also be complemented by designing algorithms that will move users away from harmful content and will amplify factual and moderate content instead.

**Question 9:** How do you envisage the proposed services in scope of this work, and in particular their design elements as they relate to the promotion of media literacy, changing and evolving within the next 5-10 years?

Confidential? – N

Currently, social media platforms contain an abundance of fake information, racist content and abusive behaviour by users. This is particularly bad when this content is used to attack individuals belonging to a group of protected characteristics such as Jews, who have suffered a considerable growth in antisemitic abuse since the war between Israel and Hamas started in October 2023. Jewish people have been marginalised, discriminated against and have been subjected to threats online (and offline). This has been driven by other users who have been exposed to extremist content about Jews and Israel, much of rooted in disinformation, misinformation and conspiracy theories.



We would hope that promoting medial literacy will turn platforms into more inclusive spaces, where threats and abused are not tolerated, and where extremist, racist and other harmful content is not easily disseminated.