Which?
2 Marylebone Road
London NW1 4DF
020 7770 7000
which.co.uk

Ofcom Online Safety Team
Riverside House
2a Southwark Bridge Road
London
SE1 9HA

Consultation Response

Which? response to Ofcom's consultation on Online Safety - Additional Safety Measures

Submission date: 20/10/2025

Summary

Which? welcomes this opportunity to respond to Ofcom's consultation on its proposed new measures which seek to build on the illegal harms and children's safety codes of practice. Our response focuses solely on measures pertaining to the illegal harms codes of practice that are designed to tackle the fraud and financial services offences category of priority illegal harm, as defined by the Online Safety Act 2023.

While we largely support Ofcom's proposals, there are a few areas where we would like to see the regulator go further:

- We believe that services should have to use proactive technology to scan **all existing content**, not just new content, for signs of fraud;
- We believe that **providers should design their recommender systems to exclude fraud content from their users' recommender feeds, as they currently have to do for other kinds of harm;** and
- We believe that **those found to have generated content judged to constitute a fraud or financial services offence should have their account removed and should be prevented from returning to the service on which they committed the offence.**

Introduction

Fraud is already the UK's single largest crime by volume, and the latest data from the Office for National Statistics, published in July 2025, suggests that the problem may be growing, with 4.2 million incidents recorded in the year to March 2025. This is the **highest estimate for this type of crime since fraud was first measured** in 2016/17 in the ONS Crime Survey for England and Wales, and it represents a **31% year-on-year increase**.

Fraud has a devastating impact on UK consumers. In 2024, fraudsters stole £1.17 billion in total from UK consumers, according to UK Finance. The damage is not only financial. Which? research has found that being a scam victim is associated with significantly lower levels of life satisfaction which we estimate to be equivalent to £2,509 per victim on average, or £9.3 billion across all fraud victims. 33% of UK consumers reported feeling anxious when using online platforms because of scams, while 35% of consumers said that they feel overwhelmed by online adverts because it is hard to tell what is real and what is fake. In extreme cases being a fraud victim can lead to physical harm: 9% of fraud victims polled by the Social Market Foundation reported an impact on their physical health as a result of fraud victimisation.

A lot of this fraud is facilitated by online platforms. UK Finance estimates that 70% of authorised fraud cases are enabled by online sources, while the Payment Systems Regulator found that 76% of authorised push payment scams in 2023 originated online. It is for this reason that Which? campaigned for the inclusion of fraud as a priority illegal harm under the Online Safety Act.

When Ofcom presented its draft Illegal Harms Codes of Practice, we warned that the proposed measures would not likely result in systemic change, because the proposals for tackling fraud applied solely to large, multi-risk services whose existing practices already exceeded Ofcom's suggested measures. We called on Ofcom to expand the scope of services who would need to deploy enhanced measures to prevent and detect fraud, and to consider the role that novel technologies could play in tackling fraud. We are pleased therefore that Ofcom has proposed new measures which seek to build on the illegal harms codes of practice, especially in relation to the deployment of proactive technologies to detect, inter alia, fraudulent content.

This response will address three areas:

- The role that **automated tools** can play in detecting potentially fraudulent content;
- The role that **recommender systems** play in disseminating fraudulent content; and
- Appropriate **sanctions** for users who are found to generate and share fraudulent content.

Automated tools for fraud detection

Question 11: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

Ofcom has proposed that services with more than 700,000 monthly users which are at high risk of fraud, and services with more than 7 million monthly users which are at medium or high risk of fraud must do the following: identify whether proactive technology which can detect fraud is available; assess whether it is technically feasible for them to deploy such technology; assess whether the technology meets Ofcom's proactive technology criteria; and, if they do decide to deploy proactive technology, use it to scan new content on their platform and remove fraudulent content where it is found. Ofcom proposes that this measure should apply to user-to-user platforms only.

We support Ofcom's proposals for two primary reasons. Firstly, we believe the harm that fraud causes to UK consumers and UK society at large necessitates an enhanced response from the regulator. We have outlined those harms in the introduction to this response. Secondly, the sheer volume of fraudulent content online and the ineffectiveness of manual, reactive reporting approaches means that proactive technology is vital to achieve a meaningful reduction in fraud content.

Online fraud is extremely widespread. A survey conducted by Yonder on behalf of Which? in November 2024 found that 22% of consumers believed that they had come across suspicious ads or messages **every single day when they were online during the previous six months**. Ofcom's own data suggests that 87% of UK adults have come across content they suspected to be a scam or a fraud. Out of 185,733 successful authorised push payment scams in 2024, 130,013 are believed to have originated online.

The evidence also suggests that a reactive approach to online fraud is ineffective. Fraud victim reporting rates are very low: the Crime Survey for England and Wales data from the year ending 2023 indicated that only about 13% of fraud cases were reported to Action Fraud (the national reporting database) or to the police. In practice, victims often report fraud to their bank, since it is the bank who will be responsible for reimbursing the victim (if appropriate). However, reimbursement of consumers is outside the scope of the OSA, and cross-sector data sharing between financial services institutions and online platforms is not happening at scale, meaning that the content which led to the victim being defrauded in the first place may not be reported to the platform which hosted it, and could be used to defraud further victims.

The harm which fraudulent content can do, and the ineffectiveness of a reporting-based approach, necessitate a proactive approach to detecting potentially fraudulent content. Our research demonstrates that proactive technologies can be effective in identifying potentially fraudulent content. In 2023, Which?, in association with Demos Consulting, collected and analysed over 6,300 adverts from Meta's Ad Library. A team of researchers then analysed 1,064 of these adverts and labelled them for whether they met a number of 'risk flags', for example claims that returns were 'guaranteed.' The team then tested whether these risk

flags could be applied automatically by training a series of neural nets on the previously coded ad dataset and using those neural nets to classify a larger dataset of 6,357 adverts.

They found that the approach was extremely effective in surfacing adverts likely to be risky. When humans alone analysed a sample of 1,319 adverts, 43% of the adverts raised at least one flag. The automated system processed 6,357 adverts, applying at least one risk flag to 186. Human reviewers reviewed 100 of the flagged adverts and found that the automated system was extremely effective in surfacing adverts likely to be risky, with only 9% of adverts incorrectly flagged. If the approach used by our team were applied by a platform in scope of Ofcom's proposals, we believe it would likely flag risky content to human moderators, allowing them to reject such content before it is published.

Given the substantial harm caused by fraud, the ineffectiveness of a reactive, reporting-based approach, and the demonstrable effectiveness of using proactive technology to detect potentially fraudulent content, **we believe that online platforms should be using proactive technology to detect fraudulent content before it is posted or as soon as possible after, in line with Ofcom's proposals.** In the case of fraud specifically, services should proactively check content against resources such as the national fraud database and the FCA's financial services register, in addition to applying Ofcom's principles-based approach.

We have one suggested alteration to Ofcom's proposals. Currently, Ofcom is proposing that services scan new content where they consider it appropriate and encourages them to do so. We would like to see this guidance strengthened, so that it becomes mandatory for in-scope services to scan all existing content as well as new content, as is already the case for child sexual abuse material (CSAM). Our research has shown that scam listings can stay active for days at a time, if not longer, meaning that scanning solely new content is likely to leave consumers exposed to huge amounts of pre-existing scam content, all of which has the potential to cause them financial and emotional damage. In order to avoid this, in scope services should be required to scan all content for potential fraud, not simply new content. As Ofcom notes, although the costs of such an activity for services with high user engagement and more frequent posting of content may be high, these services will often be able to take advantage of discounted pricing for higher volumes of content scanning (if they are procuring content scanning technology externally) and will also often enjoy higher revenues which make them more able to manage these costs. Moreover, as noted by Ofcom, scanning pre-existing content could improve the capability of proactive technologies to contextualise 'new' content, thereby increasing the accuracy and effectiveness of the technology. When scanning pre-existing content, we believe that companies should do so in reverse chronological order in order to eliminate the most immediate threats first.

Question 12: Do you have any comments on the Proactive Technology Draft Guidance?

N/A

Question 13: Do you agree with the harms currently in scope of these measures? Are there any additional harms that these measures should capture? Please provide the underlying arguments and evidence that support your views, including evidence regarding the availability of accurate and effective proactive technology.

We are pleased to see that fraud is one of the harms in scope of these measures. This is due to the serious harm fraud causes and its pervasiveness on online platforms. We have provided evidence to support this view in both the introduction and in our response to question 11, above. We have also provided evidence which speaks to the availability of accurate and effective proactive technology in our response to question 11.

Question 14: Do you agree with who we propose should implement these measures? Are there any other services that should be captured for some or all of the relevant harms?

Ofcom has proposed that services with more than 700,000 monthly users which are at a high risk of fraud (amongst other illegal harms) and services with more than 7 million monthly users which are medium or high risk of fraud should be in scope of its proposals relating to the deployment of proactive technology. **Which? supports this proposal.**

In our response to Ofcom's consultation on the Illegal Harms Codes of Practice, we pointed out that fraud was prevalent on some services which had fewer than 7 million monthly users, most notably dating services. Romance fraud is one of the fastest growing types of fraud: in April of this year, Santander reported that romance scams were one of the top rising scams in the first quarter of 2025. Moreover, data from the finance industry suggests that the number of payments per romance fraud is increasing, effectively doubling between 2022 and 2023. Despite the harm that this kind of fraud can cause, no individual dating platform appears to have sufficient user numbers to qualify as a 'large' service, using the 7 million monthly user threshold.

Which? evidence also shows that fraud is a major issue on second hand marketplaces. A Which? survey revealed at 32% of second hand marketplace buyers and 22% of second hand marketplace sellers had experienced a scam on a second hand marketplace in the two years to January 2024. The marketplaces mentioned by survey respondents included Depop, Shpock, Preloved, Nextdoor, Amazon Marketplace, Gumtree, eBay, Facebook Marketplace, and Vinted. Some of these marketplaces may have seven million or more monthly users, but not all will, emphasising the importance of also including high risk medium sized services in scope of these rules.

We therefore support Ofcom's decision to extend duties to deploy proactive technology to detect fraud to companies with 700,000 or more monthly users which are at a high risk of fraud.

Question 15: Do you agree with our assessment of the impacts (including costs) associated with this proposal? Please provide any relevant evidence which supports your position.

While the deployment of proactive technology may entail costs for businesses, these are far outweighed by the potential benefits of preventing fraud for consumers and for the wider economy.

Ofcom estimates that annual costs for the deployment of proactive technology for a larger service (7 million or more monthly users) might reach £260,000. This is in addition to one-off implementation costs, which could include £90,000 for identifying relevant third-party software. Companies which decide to build their own proactive technology, as opposed to procuring it externally, might encounter costs of £300,000 (on Ofcom's estimates) to acquire a high quality training dataset, as well as annual maintenance costs of up to £775,000.

By contrast, the UK Government's enactment impact assessment for the Online Safety Act estimates the economic and social cost of fraud to be £18.9 billion over a ten year period. Therefore, even assuming that some firms would incur costs in the tens of millions per year in order to deploy proactive technology, this proposed measure would likely still provide a net benefit even if it only to led to a modest reduction in the amount of fraud present on in-scope services.

Ofcom also suggests that the proposed measure may impact upon freedom of expression. In light of the significant public interest involved in preventing fraud and scams, we believe that appropriate restrictions on suspected content or suspected accounts would be a justified limit on freedom of expression. We agree with Ofcom's contention that there is a substantial public interest in reducing the online prevalence and dissemination of illegal content, particularly fraudulent material, which justifies the proposed intervention. The need to identify fraudulent content while protecting freedom of expression might necessitate investments in the software, skills and processes to deliver robust and accurate identification of illegal content while leaving posts expressing genuinely held opinion unaffected.

Recommender systems

Question 31: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

Ofcom is proposing that providers should design and operate their recommender systems to ensure that hate, terrorism, suicide, or foreign interference offences content is excluded from the recommender feeds of users. This should be implemented by providers of user-to-user services that have a content recommender system and are medium or high risk for at least one kind of the aforementioned harms.

Which? has encountered numerous examples of consumers being recommended fraudulent content on in-scope websites. For instance, our Scam Alerts service has picked

up fake posts on Facebook which claim to offer discounts on railcards or offers on suncream. In both cases, these fake posts direct users to malicious websites that seek to steal consumers' information. Therefore, we believe that this measure should be expanded to include fraud and financial services offences content.

We understand that the primary purpose of this duty is to prevent content from 'going viral.' We have seen evidence of fraudulent content doing just that. For instance, Financial Times journalist Martin Wolf outlined in an article from earlier in 2025 his experience of trying to get deepfake advertisements which depicted him advertising an investment WhatsApp group taken down from Meta platforms. When Wolf notified Meta of the deepfakes on 12 March, they had been seen by 106,627 users. Between 8 April and 11 April, the number of users who had seen the deepfakes increased from 222,280 to 623,780. By 24 April, the deepfakes had been seen by almost one million users. The potential for fraudulent content to achieve this kind of exponential growth in visibility necessitates platforms to prevent their recommender systems from recommending potentially fraudulent content, in our opinion.

During our response to Ofcom's call for evidence on the first phase of online safety regulation, we also highlighted the issue of boosted content. Fraudsters can use boosted content to enhance the reach of their scams without being subject to the same scrutiny that those creating fraudulent advertisements will be, since boosted content still counts as user-generated content under the Act. We urge Ofcom to revisit boosted content within these Codes of Practice after publishing the Code of Practice for Fraudulent Advertising to ensure that the same protections that apply to fraudulent advertising apply to boosted content.

Question 32: Do you have evidence on what types of content are typically recommended to users as part of concerted foreign interference activity?

N/A

Question 33: Do you have evidence on whether services track the extent of algorithmic amplification, such as impressions and reach, of content that is later deemed illegal/violating. If so, do they (or does your service) use this information to enhance the safety of their systems?

N/A

Question 34: Do you agree with our assessment of the impacts (including costs) associated with this proposal? Please provide any relevant evidence which supports your position.

Ofcom has assessed that service providers will face direct costs associated with introducing the measure, as well as potential indirect costs if they have a business model which generates revenue in proportion to levels of engagement with certain kinds of content. In both cases, Ofcom considers that the benefits of preventing UK internet users from being recommended in-scope content justify the measure. Given the previously articulated economic and wellbeing costs associated with fraud, we agree with Ofcom's contention.

Ofcom has also assessed that the proposed measure has the potential to interfere with users' rights to freedom of expression, freedom of association, and to privacy. Ofcom considers that the potential interference with users' rights is proportionate in light of the benefits that protecting users from relevant harms would secure. We agree with Ofcom's assessment, for reasons already cited in our response to question 15.

Question 35: Are there any impacts of the proposed measure that we have not identified? Please provide the rationale and any supporting evidence for your response.

N/A.

User sanctions

Question 39: Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

Ofcom is proposing that providers should prepare and apply a sanctions policy in respect of UK users who generate, upload, or share illegal content and/or illegal content proxies, with the objective of preventing future dissemination of illegal content. Ofcom is also proposing that providers should set and record performance targets for their content moderation function covering the time period for taking relevant content moderation action.

Which? agrees that users who have been found to have uploaded fraudulent content should be sanctioned by the platform to which they uploaded the scam content. Ofcom has suggested that a "strikes" system, wherein offenders are given warnings for initial infringements, could be appropriate.

We believe that, when it comes to fraud, the rules should make a distinction between users who generate fraud content, and users who share or upload fraud content. Ofcom's Illegal Content Judgements Guidance states that "it will always be difficult to identify fraudulent content because fraudsters use tactics to appear legitimate." As such, it is possible to imagine users sharing or uploading fraudulent content which they believed was genuine. In such a circumstance, a warning from the platform would be an appropriate sanction.

However, we do not believe that a warning is an effective or appropriate sanction in the case of users who generate fraudulent content. To make an illegal content judgement in relation to fraud or financial services offence, a service must be able to show that the user has committed one of the relevant offences listed in Schedule 7 to the Act, which in turn requires the service to be able to make an inference of dishonesty or false representation on the part of the user who generated the relevant content. Ofcom's own guidance suggests that the threshold for making such a judgement is likely to be "difficult to reach." The way the Act works means that services are only likely to make content judgements regarding fraud offences when they are fairly certain that the user who generated the content had serious intent to defraud or knowingly represented themselves as something they were not. Therefore, in situations where the user has generated the content

themselves, or knowingly assisted the person who did so, we feel that the appropriate sanction should be an instant revocation of the relevant user's account and a ban from the platform thereafter. Such a sanction is already in place for users found to have uploaded, shared, generated, or received child sexual exploitation and abuse content.

Our research shows that the majority of fraudulent advertising is carried out primarily by organised criminal groups who are financially motivated, and that even highly capable individuals who can carry out attacks are still reliant on infrastructure and resources established by other parties in the criminal chain. Likewise, a recent investigation conducted by the Organised Crime and Corruption Reporting Project detailed the sophisticated infrastructure underpinning many scam operations. Which? has seen reports of scammers paying for independent reports on products they have listed on online platforms, as well as setting up fake companies and providing fraudulent verification details such as equipment register numbers. We do not believe that warnings will be effective for people whose full-time job is to defraud consumers.

Of course, one can argue that professional fraudsters who have their accounts deactivated will simply create new accounts with new contact details, and this is likely true, but removing their accounts will at least disrupt a fraudster's operation for a period of time and may reduce losses overall, while warnings will achieve nothing whatsoever. Moreover, Ofcom has given services an obligation to prevent those who upload child sexual exploitation and abuse content from returning to services after having been removed, and has left the technical means of doing so in the hands of the services. There is no reason that such an obligation could not also be in place in relation to those found to be generating fraud content.

To summarise, given the intentional and highly professional nature of online fraud, we believe that services should revoke the accounts of those found to have generated content which constitutes a relevant fraud or financial services offence under Schedule 7 to the Act and prevent those users from returning to the platform. We believe that a system of warnings remains appropriate for those found to have shared or uploaded content which amounts to a relevant fraud or financial services offence, since, although it could be deliberate, it is possible to imagine users doing so accidentally.

We also believe that the user sanctions measure with respect to fraud should apply to all users, not just UK users, as is already the case for child sexual abuse material. Fraud is a highly international crime and fraudsters use the internet and telecoms networks to attack UK consumers from countries all around the world. For instance, the previously-cited report from the Organised Crime and Corruption Reporting Project details investment scams originating in a range of geographies, with Georgia and Israel being two of the most prominent examples. The City of London Police estimates that around 70% of fraud offences in the UK have ties to overseas criminals, with around £8.2 million lost each day to overseas accounts in 2023. Given this reality, only sanctioning UK users is unlikely to address the full extent of the problem. We believe, therefore, that the proposal should be extended to cover international users too.

With respect to performance targets, given that the targets are set by services themselves and not by Ofcom, over time we would expect to see industry benchmarking of performance and platforms failing to meet these standards held to account by the regulator.

Question 40: Do you agree with our assessment of the impacts (including costs) associated with this proposal? Please provide any relevant evidence which supports your position.

Ofcom expects that service providers may incur one-off upfront costs as well as annual ongoing costs, which will vary significantly across services. Larger services are likely to incur higher costs, but may also be in a position to automate their user sanctions policy, which could lead to cost savings. As we have stated in previous sections, we feel that any costs incurred are likely to be enormously outweighed by the cost of doing nothing, which the government estimates at £18.9 billion over the next ten years.

In terms of rights, Ofcom anticipates that its proposed user sanctions measures will have implications for the freedom of expression, association and the right to privacy. We agree with Ofcom's contention that its intervention is justified and proportionate "in light of the risk of harm to users posed by illegal content." Ofcom's guidance obliges providers to operate a complaints procedure which gives users the right to appeal against a user sanction applied to them. This should ensure that legitimate expression is protected, whilst illegal content is removed from services.

About Which?

Which? is the UK's consumer champion, here to make life simpler, fairer and safer for everyone. Our research gets to the heart of consumer issues, our advice is impartial, and our rigorous product tests lead to expert recommendations. We're the independent consumer voice that works with politicians and lawmakers, investigates, holds businesses to account and makes change happen. As an organisation we're not for profit and all for making consumers more powerful.

For more information contact:
Matthew Niblett
Senior Policy Advisor
matt.niblett@which.co.uk

October 2025