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 protecting people from illegal harms online

Submission date: 23/02/2024

Summary

Which? welcomes this opportunity to provide our views on the proposed requirements to be placed on online services to protect UK consumers from illegal harms. Our response focuses on fraud detection measures proposed in the draft Illegal content Codes of Practice and Guidance.

While we recognise that Ofcom's proposals are a step in the right direction, in protecting consumers online, we do not anticipate the measures proposed for fraud detection will result in systemic change. We note that Ofcom's fraud detection proposals have been assigned to large online and multi risk services, yet these services already have systems and practices that go above and beyond Ofcom's proposed measures.

In order to leverage opportunities presented in the draft Codes of Practice and Guidance to provide consumers with greater protections against fraudulent online content, we propose that Ofcom consider the following measures:

- **Calculation of Online Service Size** - we recommend Ofcom consider applying the following options with the intention to capture online services such as dating sites where fraud is prevalent.
    - 1a: Extend fraud detection duties to services with over 700,000 users
    - 1b: Broaden the scope of risk profiles to include social context of online service
    - 1c: Calculate large online services through its group capacity, revenue and reach

- **Use of URLs to detect fraudulent content**
    - Services with over 700,000 users should be required to use URL detection to

detect possible fraud
    - Add using unknown contact points as a risk factor in the Illegal Content Judgement Guidance

- **Future proofing fraud detection**
    - Ofcom should use its information gathering powers to generate an evidence base on the use of new technologies (including artificial intelligence) to tackle fraud.
    - Ofcom should create a standard in their codes of practice that online services use specific types of data to train their systems to prevent fraud
    - Add Specified Anti Fraud Organisations (SAFOs) to the list of trusted flaggers

- **Future considerations**
    - Apply safety by design principles to all verification schemes
    - Consider the risk of fraudsters exploiting boosted content

Full response

The harm landscape

When victims are scammed, it is hardly ever confined to a single service. Which? research has exposed how organised crime gangs use scaled operations across multiple channels. Victims are moved from one digital online service to another until they reach their ultimate goal of payment. Fraudsters are confident that with the volume of reach, they will be able to get to enough victims who are in a vulnerable state, to make incredible amounts of money.

Calculation of Online Service Size

We note that Ofcom have recommended the most impactful measures in protecting UK consumers from illegal harms for large and multi-risk services only. In the draft illegal content Codes of Practice, Ofcom has proposed to define a large online service as one with over 7 million UK users per month (11.51). While Which? recognises the need to target services with the widest reach, our evidence suggests that online services with high risk of fraud such as online dating apps will be excluded from these obligations.

Which? believes that Ofcom should consider alternative criteria or thresholds for services that have a high risk of fraud but have less than 7 million UK users per month.

Recommendation 1a: ***Extend fraud detection duties to services with over 700,000 users***
Which? notes that Ofcom has included recommendations for smaller online services that has been proposed for large online services in relation to automated tools for the detection of Child Sexual Abuse Material (CSAM) (4G, 4H). This extends the obligations to

services with over 700,000 UK users per month.

Building on the CSAM proposals, Which? recommends that a similar distinction is made for fraud detection. Ofcom already recognises that online services with 700,000 to 7 million UK users per month have the potential to impact a significant number of users if harm is present on their service. In its risk assessment guidance (table 6), Ofcom states that these services should be considered medium risk. This ensures that these services with a high risk of fraud are not overlooked in the regulatory framework.

By extending certain fraud obligations (such as direct reporting channels to trusted flaggers) to services with over 700,000 monthly UK users, Ofcom can ensure that these services are subject to obligations similar to those placed on larger services, without imposing an undue burden on the smallest services.

Recommendation 1b: ***Broaden the scope of risk profiles to include social context of online service***

Online dating services operate in a distinct social context that could make the service vulnerable to various forms of harm, particularly fraud and grooming. The nature of these services, characterised by personal interactions and the exchange of sensitive information, present challenges that are not addressed in the proposed risk profiles. Which? believes that the scope of the risk profiles should be broadened to include how the purpose of a service can make it vulnerable to harm.

Our investigations show that fraudsters push for an emotional connection quickly and use this connection to groom their victims over months or even years into believing that the connection is genuine before requesting money and bank/ credit card details. Fraudsters are able to manipulate our natural, human biases to build trust and overwhelm the rational mind into believing that an interaction is authentic.

Ofcom's current proposals for the risk profiles do not address the social context in which these services operate. Which? recommends that Ofcom consider the wider social aspects of a service in accordance with its risk of harm as part of its risk profile. Recognising and addressing the wider aspects of online services can contribute to building greater online protections for consumers against multiple illegal harms and boost user confidence.

Recommendation 1c: ***Calculate large online services through its group capacity, revenue and reach***

Ofcom has assigned the most onerous measures to online services that they have determined to be large according to their reach. It has justified this on the basis that they have the financial capacity to absorb costs (11.53-60). The proposed approach does not consider the capability of online services that fall below the current 7 million users per month that have the financial means by which to undertake these measures.

Ofcom's judgements on size are based on a single service rather than the capacity of the organisation. Ofcom's proposed approach could incentivise organisations to have multiple smaller services rather than a single larger service. For example, Match Group operates 12 different online dating services with an annual revenue of $3.4bn, none of the services individually appear to have sufficient user numbers to qualify as a large service.

Ofcom considered defining large services based on annual revenue. It suggested that this may not always accurately capture access to resources. Which? does not believe that the calculation of size based on user numbers accurately captures access to resources either.

We recommend that Ofcom apply its considered revenue threshold £10 - £50 million at a group level in addition to individual online services that are considered high risk of fraud. In the event that a single service does not meet this threshold, their group revenue should be considered and this will ensure that sectors such as the online dating community where fraud is prevalent are brought into scope.

Our suggested aims to establish a duty on smaller services to actively protect consumers from fraud, acknowledging each one has unique challenges. Which? believes smaller services also have a responsibility to protect their customers from fraud. Our recommendations aim to strike a balance between the level of harm and the feasibility for these services to implement fraud detection measures currently proposed in the Codes of Practice.

These proposals encourage a safety by design approach to be mindful of the increased threat from bad actors as they grow in users. It caters to the diverse needs of services at different stages of development, fostering innovation while safeguarding consumer interests.

Evidence of level of harm on online dating services

Ofcom's analysis of UK adult users of online dating apps, in its most recent Online Nation report, indicates that the most widely used online dating app in the UK has under 2.5 million users per month. These services are undeniably significant within the online user-to-user landscape and have a high risk of fraud. We are concerned there is a risk to users that, under Ofcom's current criteria for online large services, these types of services have no extensive content moderation duties.

Online dating romance fraud is becoming an increasingly common experience for many in the UK. There was a rise in romance scams in 2022 with £31 million stolen. Which? investigations have revealed countless experiences of consumers who have fallen victim to

romance fraud through online dating apps. Fraudsters will rely on emotive tactics to build a trusted persona with their victim and will ask for money, bank details or gifts when the victim's defences are down. Our evidence highlights that romance fraud reached a staggering £73.9 million in the year to April 2021, with 7,754 reports filed to Action Fraud.

Although the figures of this harm are staggering, it likely underestimates the true extent of this issue. Victims of romance fraud often find themselves too embarrassed or emotionally distressed to report incidents to authorities, leading to a substantial gap in the reported numbers. This underreporting not only skews the statistical landscape but also makes it difficult to fully understand the magnitude of the number of UK users affected by this.

Use of URLs to detect fraudulent content

Automated detection of URLs

Proactive measures (i.e. automated detection to identify fraud before it can reach consumers) are a key element in preventing consumer harm. There are indicators that content is fraudulent that can be reliably detected by automated systems. Online content can include internet locator information (URLs, IP addresses or blockchain domains) that link to fraudulent websites or services.

Recommendation: ***Services should be required to use URL detection to detect possible fraud***
Suspicious URLs are a useful piece of data which can indicate potentially fraudulent content. However, the presence alone of a suspicious URL is not enough to provide reasonable grounds to infer that the piece of content is itself fraudulent. Which? believes that services should be required to use relevant data such as URLs to proactively detect possible fraud on their services. Combining URL with other sources of data presents a picture of whether content is fraudulent. As a result we do not believe that URL matches should lead to automatic takedown but instead should require that the content is processed through the organisation's content moderation system (as in A4.46). We recommend that services should be responsible for procuring an accurate feed of likely fraudulent URLs and other internet locator indicators. This should include having established mechanisms for reviewing the accuracy of the locator information provided to ensure low levels of false positives (as in A4.53-A4.57).

Given the low cost of the cheaper options available, fraudulent URL detection should not just be a requirement for the largest services. Ofcom should consider recommending Fraudulent URL detection for services with over 700,000 users as with CSAM URL detection (A4.35).

Evidence of available accurate URL data sources

We understand from Ofcom's recommendation to use URL matching to detect CSAM that there are no issues with the technology itself. Ofcom's concerns are focused on the sources of URLs. This includes a source's *completeness, accuracy and risks of bias* (14.169-170). Which? has identified a number of ways to ensure that services acquire accurate URLs to overcome the risks Ofcom have outlined in the consultation. This includes internal processes, Government provided URLs or private feeds of URLs.

Internal

Services can use URLs that their internal processes, such as user reporting, have identified as fraudulent. In some circumstances, where services remove content that they have identified as fraudulent they should also store fraudulent URLs and use this to detect new content that is likely to be fraudulent. For example where a service identifies a post impersonating a brand or well known organisation

Government Provided

The National Cyber Security Centre (NCSC) operates a "Share and Defend" hub. The NCSC collects suspicious URLs and domains and provides them to Internet Service Providers (ISPs). ISPs check the URLs they receive and block from their channels those they agree are suspicious. The NCSC is planning to expand this functionality to tech companies.

Our engagement with ISPs has revealed that this information has been helpful for finding and blocking fraudulent URLs. We cannot provide direct evidence of this as the NCSC does not publish accuracy rates for blocking based on its service. However, this information should be available to Ofcom from NCSC and the ISPs it partners with.

Private feeds

There are a large number of commercially available feeds that supply suspicious URLs. As the DNS Research Federation highlights in [CONFIDENTIAL] these can be synthesised into actionable intelligence to remove fraud. Services can generate a more accurate and less biassed list of fraudulent URLs for use by combining a number of different external feeds of suspicious URLs alongside data from Passive DNS monitoring. There is little overlap between the URLs flagged on commercial feeds, so a combination of sources leads to more accurate information. Using multiple URL feeds and additional data points can help to filter out results reported as spam rather than fraud and limit the use of outdated information.

Combined systems can also help build a feedback loop that can continuously improve the quality of existing feeds by establishing accuracy rates. DNS Research Federation's DAP.LIVE platform is currently used by a large online service for its enforcement activity to identify brand impersonation and is used to tackle 19,000+ URL threats mimicking their domains each year.

Costs

We are confident the costs associated with procuring this data is affordable for services in scope of the current recommendation and for those we are recommending be included.

Our discussions with key stakeholders has shown that URL sharing most commonly relies on the same technology that these services use to share cybersecurity threats. For example we understand that the Share and Defend programme currently shares data through the Malware Information Sharing Platform (MISP) or Trusted Automated Exchange of Intelligence Information (TAXII) standards. Major service providers including Google and Microsoft are key participants in setting the TAXII standard. Google offers MISP integration into its services. Meta operates its own ThreatExchange platform for sharing threat intelligence which includes domains.

Evidence from the DNS Research Federation in [CONFIDENTIAL] shows that costs for participating in their data acquisition service are at the lower end of the scale anticipated by Ofcom. Large businesses can access a full set of features for [Confidential] per year whilst small businesses could have a more limited set of features from [Confidential] per year. These features can include access to APIs for bulk analysis of newly detected fraudulent URLs.

Services with over 700,000 users are already required to use URL matching technology in order to combat CSAM. This means there would likely be minimal extra development costs.

URLs in the Illegal Content Judgement Guidance

Which? believes that the Illegal Content Judgement Guidance (ICJG) can be more effective by being more specific about data that will help services detect fraudulent behaviour. Which? supports the approach set out in the ICJG to identify fraud by false representation using three groups of criteria (A6.37-6.42). However, it does not include any detail on how to judge whether a well known organisation or brand is being impersonated.

Recommendation: ***Add using unknown contact points as a risk factor in the Illegal Content Judgement Guidance***

Currently the ICJG includes the **"use of an account or page which claims to represent a public figure, well known organisation or brand, unless it is obviously run as a parody."** This guidance could be made more useful for services by specifying indicators that content or an account is not genuine. We recommend that the following is added to Group 1 on disguised account information or activity:

An account which claims to represent a well known organisation or brand provides or links to a contact method (for example, a website, telephone number or email address) different from that brand or organisation's known official channels.

Evidence of harm from brand impersonation using URLs

Brand impersonation is a common tactic used by fraudsters to manipulate online users. Which? notes that while Ofcom recognises the prevalence and harms associated with brand impersonation online (6O.10, 6O.31, 6O.33, 6O.63-65), it does not connect this to the use of links to fraudulent sites.

Brand impersonation can be part of a sophisticated attempt to create a fake identity online that will draw in potential victims through social media to a fake website. The Crime Survey for England and Wales shows that brand impersonation is a common element of fraud with delivery companies, banks and ecommerce companies most likely to be impersonated. The survey for the year ending March 2022 found that 1.6% of adults in England and Wales (equivalent to around 760,000 people) clicked a link in an impersonated message on email, text or social media in the last year. Which?'s latest research has found fraudsters pretending to be major retail brands linkin from social media to fraudulent websites.

The North West Cyber Resilience Centre, a joint venture between North West Police forces and Manchester digital, highlights that fraudsters can use a new URL that is only subtly different from that of a genuine brand in a way that is hard for consumers to spot. For example a scammer could use a look-a-like domain like 0fcom.org.uk. As the Government launched its new fraud advice website www.stopthinkfraud.campaign.gov.uk; Which? was able to buy www.stopthinkfraud.co.uk without any resistance. This demonstrates the ease at which copycat domains can be purchased and potentially used for malicious purposes.

Research from the DNS Research Federation, finds that when consumers are shown a URL the presence of a brand name anywhere in the domain or subdomain increases trust and can make someone believe that it is legitimate. For example, a consumer may believe that ofcom.onlinesafety.net is a legitimate Ofcom site despite this being a purchasable domain[1]

[1] As of 19/1/24 onlinesafety.net with the sub domain ofcom.onlinesafety.net was available to purchase.

unconnected to Ofcom. In accordance with our recommendations on matching URLs and the ICJG guidance, if a service notices that a URL is different to the known one for that organisation (as with the Ofcom examples above) this is an indicator it could be fraudulent. Equally, if the URL in content has been previously used fraudulently, then a service using URL detection would highlight the content to its moderation systems and remove it if it were fraudulent.

Domains registered through the centralised Domain Name System (DNS) only allow for a single holder of a domain (for example there is only one Ofcom.org.uk) whilst with a blockchain domain there can be duplication. Each blockchain-based domain may have duplicate names within another blockchain (for example a different person could own Ofcom.wallet on the Handshake service to the person that owns Ofcom.wallet on the Unstoppable Domains service). This is called a domain name collision. Fraudsters can take advantage of domain name collisions to trick users.

Blockchain domains are harder to takedown than standard domains. In the DNS system Top Level Domains (TLDs) are overseen by registries (for example Nominet oversees the .uk TLD) and can remove bad actors using their domains. As blockchains domains are decentralised there is no equivalent ability. This means that a blockchain domain used in fraud could persist for longer and can most easily be addressed through other services blocking or removing links to access it as per our recommendation URL detection on matching.

Additional research from the DNS Research Federation has highlighted that major brands have registered some blockchain domains and that there are risks of blockchain domains being used in phishing. This could become a future avenue for scammers to use in brand impersonation. Services monitoring contact points would be in a better position to take action to remove content with fraudulent blockchain domains as with our recommendation on the ICJG using available data sources.

Mobile network operators have found some success in preventing brand impersonation scams by establishing checks that a contact point is genuinely associated with a brand. The Mobile Ecosystem Forum has an SMS Sender ID Protection Registry. This allows organisations to register the methods they use for sending SMS with their brand name to help telecoms networks more easily spot attempts to impersonate them. Online services can similarly compare known contact points (like URLs, blockchain domains, IP addresses or telephone numbers) with ones included in content and could over time create an equivalent registry.

Future Proofing

Machine learning

The Codes of Practice risk being outdated before they are in effect. Ofcom has judged that there is not yet a sufficient evidence base on the deployment of technologies such as machine learning (ML) or artificial intelligence (AI) to detect previously unknown content (14.322-14.328).

Recommendation: ***Ofcom should use its information gathering powers to generate an evidence base on the use of new technologies (including artificial intelligence) to tackle fraud.***
Ofcom should, as an immediate priority, use its newly granted information gathering powers to rapidly generate an evidence base on the use of machine learning to detect illegal content including fraud. Which? recognises the lack of publicly available data on the accuracy of service's ML systems. The absence of clear recommendations encouraging the responsible use of this type of technology, risks allowing sophisticated organised crime to continue to be one step ahead. It also risks that services are not being incentivised to invest in this technology to continuously improve their detection and take down.

A case in point of outdated technology is Ofcom's recommendation on the use of keyword matching (4I). Our data scientists and external experts we asked about keyword matching stressed that the technology was outdated and frequently manipulated or avoided by fraudsters. For example, fraudsters use "cooking" to describe creating counterfeit checks, a commonly used word that could not be added to a keyword list.

Experts recommended they would use machine learning in detection of articles in the use of fraud. Which?'s own research has found that machine learning classifiers can be trained to detect content that is likely to be fraudulent. We trained four different models to detect if:

1. content advertised returns as guaranteed;
2. promised life-changing returns;
3. created a sense of urgency
4. or played to the consumer's peace of mind.

Our models were tuned to minimise false positives this allowed us to have high rates of accuracy and precision although lower rates of recall.

Model	Accuracy	Precision	Recall

Guarantee	91%	100%	25%
Life changing	82%	57%	22%
Peace of Mind	80%	42%	17%
Urgency	77%	60%	13%

From our sample of 6,357 adverts we identified 57% which raised three or more red flags. This was not new or cutting edge technology. It was done using open source tools originally developed by Google engineers in 2015[2] and was undertaken by a small team with less resources than would be expected from a relatively small online service. We did not have access to metadata about content. We would expect that any service using the same tools to detect fraud would have higher accuracy rates due to having access to substantially more data to look for signals of possible fraud.

AI and ML are increasingly becoming standard tools of fraud prevention across different industries. The Payments Association noted that ML has been widely used for several years to detect suspicious patterns in bank transfers. Major payment and hosting providers offer off the shelf solutions. Our engagement with large online user to user and search services has suggested that they use ML in a similar way to detect fraudulent content. There is a lack of publicly available data on the accuracy rates of these efforts however these are standard practices in the industry.

A current example of AI and ML is in behavioural biometrics which is already a security feature used by banks including HSBC.
The latest developments in Generative AI have suggested that there may be new uses of AI to improve content moderation. Industry leaders are exploring the possibility of using this new technology to improve efforts to detect harmful content. Academics have found that with relatively little effort these models can be used to create content moderation systems that perform better than existing industry tools.

It is unclear how Ofcom's need for a substantial evidence base before recommending technology specific interventions will keep pace with these latest developments. There is a real risk that the technology specific approach combined with the need for a substantial evidence base will continue to lead to recommendations which do not encourage services to use the most accurate and impactful technology. As a result, more consumers will be victims of fraud. In light of the evidence requirement constraint - we recommend Ofcom use its information gathering powers to full effect - with a focus on the consumer outcome being met with the best technology available at any given time.

[2] Models in the research were trained using Keras and Tensorflow

A technology neutral data led approach

We believe the intention of reducing fraud online will be more readily achieved if Ofcom has a broader outlook on how to approach fraud detection that avoids the constraints of technology. To do so this must focus on the data which builds effective fraud detection systems.

Recommendation: **Ofcom should create a standard in their codes of practice that online services use specific types of data to train their systems to prevent fraud**
Ofcom can provide a technology neutral approach to improve content moderation systems to prevent fraud by focusing on data. Ofcom has created a framework of indicators in the ICJG, and should use this to create a standard in the codes of practice for specific data use in the prevention of fraud. Examples of types of data can be seen in Figure 1.

The seeds of this approach already exist in some of Ofcom's existing recommendations. For example, Ofcom recommends that services should track new illegal harms using relevant data from:

- its complaints processes,
- content moderation processes,
- referrals from law enforcement
- and information from trusted flaggers or other expert groups (3E).

Ofcom currently provides no direct link between the data sources based approach to risk assessment and the content moderation processes recommendations. Which? has demonstrated that effective fraud detection requires a variety of data indicators including the use of known fraudulent URLs to detect suspicious content and authentic URLs to aid in illegal content judgements. Ofcom could improve services' systems by requiring that they collect and use certain indicators in the development of their content moderation systems. Figure 1 provides examples of indicators Ofcom suggests in its guidance and what equivalent recommendations for data services should access might look like.

Figure 1:

Indicator from ICJG	Examples of Recommended data
Claims that the investment or firm concerned is regulated by a body which does not exist. (The presence of this example, in addition to one group 2 indicator will be sufficient to provide	Services should maintain access to an up to date list of the relevant regulators.

reasonable grounds to infer that content is illegal).	
Use of an account or page which claim to represent a public figure, well known organisation or brand, unless it is obviously run as a parody.	Services should maintain access to a list of major official accounts, and pages for public figures and brands.
Posts using enticing language to suggest unrealistic gains; for example, 'easy money' or 'fast cash'; posts which offer investment opportunities with a high, unrealistic rate of return within the time frame for investment or current environment, or which otherwise seem 'too good to be true'; posts which exert pressure on those being requested to send money or invest, including time pressure which is not warranted.	Services should identify a collection of example content of this type in consultation with persons with expertise in the identification of content that amounts to fraud by false representation.
user accounts which heavily feature content which contains wealth signifiers (e.g. currency, luxury cars, private aeroplanes, snapshots of bank accounts), where these appear to have been gathered from multiple sources and are being used to back up claims about investments	Services should assemble a database of example images of wealth signifiers in consultation with persons with expertise in the identification of content that amounts to fraud by false representation

A requirement to use these types of data would incentivise a model of continuous improvement to drive down fraudsters reaching consumers. The data can be used to fuel proactive technology systems and would also be useful for improving the training of human moderators.

As a result, regardless of the technology deployed, services are able to detect and tackle more consumer harm. This should be considered alongside the growth of 'fraud prevention as a service' where there are now a wide range of off the shelf packages of technological solutions. By Ofcom focusing recommendations on the data being used rather than the type of system, it allows more room for innovation in the market to meet the intended outcomes for the fraud prevention services.

Ofcom can ensure that user-to-user and search services are part of a connected counter-fraud ecosystem. Currently fraud data sharing between banks and online services is mostly in isolated pilots. These schemes have the potential to scale up to provide fully-fleshed cross sector intelligence sharing. Ofcom must encourage this to be adopted as swiftly as possible. Which? has shown that fraudsters will use all available channels to reach

a consumer and only through collaboration between sectors can fraudsters be halted from reaching consumers.

Trusted flaggers

Which? believes that trusted flaggers will result in enabling services to build more effective systems that continuously improve over time. We are pleased that Ofcom has recognised the importance of online services using data from sources within and outside their organisations to offer particular expertise (5I). This will assist online services in notifying the presence of potentially illegal content on their services.

Recommendation: **Add Specified Anti Fraud Organisations (SAFOs) to the list of trusted flaggers**
The current list of trusted flaggers misses organisations that have important data which could be useful for identifying fraud on services. The Serious Crime Act 2007 (order 2008) lists a number of Specified Anti-Fraud Organisations (SAFOs). SAFOs are trusted organisations, recognised by the UK government, to enable and facilitate the sharing of information across both the public and private sectors. Which? is aware that multiple SAFOs are keen to work with service providers to use their data to better detect fraud. Their inclusion ensures that the regulatory framework leverages the knowledge and capabilities of organisations with a proven track record in this domain.

Which? recommends that Ofcom consider the use of Application Programming Interface (API)s between trusted flaggers and online services. As these systems expand, they would allow for bulk data to be used in real time. This would allow existing trusted flaggers like the NCSC to feed in data on suspicious activity from the Share and Defend programme at scale and at speed.

Which? would like to highlight that this is a key recommendation and should be applied to the wider scope of online services as per our recommendation 1a/b/c. In the case of online dating fraud, Action Fraud provides free access to its data on romance fraud that online dating app services can find useful when moderating potentially illegal content.

Future considerations

User verification

Consumers should be protected from the harm of poorly designed verification schemes regardless of size. We welcome Ofcom's proposal to recommend that online services implement good design practices when providing user verification schemes (9C).

Recommendation: **Apply safety by design principles to all verification schemes**
Ofcom proposes limiting these obligations to only large services that label user profiles as notable or offer a monetised scheme. As Ofcom's evidence (20.83-95) suggests that the harm comes from the poor design of these schemes it is unclear why this measure should apply only to schemes on large services. There is no evidence to suggest that a verification scheme that failed to comply with Ofcom's proposals would be beneficial to users at preventing rather than facilitating impersonation. Given this risk to consumers and the lack of benefit Ofcom should apply these measures to all services that label user profiles as notable or offer a monetised verification scheme.

Fraudsters have used weaknesses in verification schemes in order to mislead victims into believing false identities are authentic. Cybersecurity experts have highlighted that it is currently easy for criminals to purchase social media verification and use that to defraud consumers. Ofcom's recommendations are a positive step in ensuring that service design prevents this type of attack.

We urge Ofcom to consider revisiting the recommendations in the Code of Practice to align them with the results of its future consultation into user empowerment. In the user empowerment consultation process for that guidance Ofcom should seek to establish an understanding of best practices for user verification.

Boosted content

Ofcom's proposals have not included any provisions to tackle the increased risk posed by user generated content that is promoted after receiving payment. The Online Safety Act 2023 definition of fraudulent advertising excludes regulated user-generated content. This means that user generated content which acts like advertising (in that it is promoted to users on the basis of payment), also known as boosted content, is covered by the Illegal Content Codes of Practice rather than the Fraudulent Advertising Codes of Practice. As a result any additional requirements placed on fraudulent advertising do not apply to boosted content. Our primary concern is that bad actors could avoid due diligence checks and still reach consumers at the same scale as paid for advertising.We believe it is essential to address this gap.

Which? has previously raised concerns about how paid for boosted content could be used to effectively promote fraudulent content to consumers. 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 prevent fraudulent advertising apply to boosted content.

Conclusion

While we appreciate the intentions of the proposed Codes of Practice, Which? believes that they do not adequately address the challenges posed by large online services. To truly achieve the intended objectives of the Online Safety Act, it is crucial that the Codes focus on outcomes rather than mere compliance. By focusing on measurable outcomes, Ofcom can ensure that the regulatory framework is effective in fostering a safer and more accountable online environment for all users. Thank you for considering our input, and we look forward to contributing further to the development of a robust regulatory framework.

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:
[✂]

[✂]

February 2024