Consultation: Protecting people from illegal harms online

Date: 23 Feb. 2024

Address: 2a Southwark Bridge Road, London, England SE1 9HA

Sent to: IHconsultation@ofcom.org.uk

UK Finance is the collective voice for the banking and finance industry.

Representing more than 300 firms across the industry, we act to enhance competitiveness, support customers and facilitate innovation.

Executive Summary

We welcome the Ofcom consultation and extensive codes of practice that have been drafted as part of the illegal harm package. Our core recommendations detail where enhancements to the proposed codes of conduct would ensure effective implementation of the Online Safety Act (OSA), the Financial Services and Markets Act (FSMA) and Proceeds of Crime Act (POCA), and result in greater prevention of fraud and the harms it inflicts on society. They are as follows:

Better utilisation of data from customers reporting fraud to financial services

- We were encouraged to see Ofcom acknowledge the scale of the fraud problem facing the UK. Over **2.9m frauds and scams were reported in 2022** according to UK Finance [1]. Financial Services Trade Bodies and Specified Anti-Fraud organisations with an intelligence function (such as UK Finance, Cyber Defence Alliance and CIFAS) ***receive invaluable data from their its membership on key threats***. As a minimum, this intelligence resource should see them included as Trusted flaggers to help all participants in the fraud ecosystem keep pace with the constantly adapting behaviour of criminal actors.
- The Office of National Statistics report said **3.2m offences were reported** in England & Wales year ending September 2023 [2]. They also note that the true figure is likely much higher as fewer than **1 in 7 crimes are reported**.
- We also recommend, ***as a minimum*** that banks and building societies with over 7mn users should be independently eligible for trusted flagger status. These firms are often the first point of contact for victims and could effectively gather the data required to get illegal content removed quickly in order to reduce the risk of further harm. These firms mirror the OSA 'large' service definition, ***have specialised knowledge and intelligence to share that comes from encountering tens of thousands of victims first hand.***
- In addition, Ofcom should leverage its information gathering powers to generate and publish an industry view of illegal harms reported because of using an online service. This information will enable consumers to assess the risks when interacting with an online service, and through benchmarking will encourage services to do more to improve their preventative controls.
- To do this Ofcom could consider using the case reports (confirmed financial losses) on Authorised Push Payment (APP) Scam, from Financial Service Institutions, to assess a Services adherence to the act. To go one step further, this comprehensive data could be

[1] https://www.ukfinance.org.uk/system/files/2023-05/Annual%20Fraud%20Report%202023_0.pdf

[2] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingseptember2023

used form the equivalent of a user complaint, giving Ofcom valuable insight to help assess whether individual companies are doing enough to protect their users from illegal harms.

- Similar benchmarking data is produced by the Payments Systems Regulator to assess the effectiveness of the financial services industry's controls. Ofcom could also look to assess the speed with which online services remove content when a trusted flagger reports a concern with content.

Increased requirements for Services to stop criminals operating on services

- **Fraud accounts for 40% of reported crime** in England and Wales [3], and the majority of scams begin via online services. As such fraud should be in scope as a trigger for automated content moderation and user access restrictions, as opposed to only applicable for CSAM or terrorism. Fraudulent funds are often seed money for other organised crime.
- Online Services that facilitate high levels of fraud due to user anonymity or pseudonymity such as social media, online marketplaces and dating sites need to have Identity and Verification (ID&V) requirements in place as a minimum. This would significantly limit a range of harms including impersonation, purchase and romance scams by making it easier to identify perpetrators of harm and, most importantly limiting their entry onto Services, through fake profiles, in the first place.
- Greater use of proactive technology is required to strengthen effectiveness (including but not limited to AI, machine learning, analysis of linguistic and stylistic cues, image crawlers), as well as real time information and intelligence sharing both within and across sectors. Many of the large online services are experts in this field so should be well placed to introduce these controls by design.
- Ofcom should require online service providers to educate their customers about the risks of fraud and scams on their online services. This should include both customer communications (like Take Five for the FS sector), but also effective and timely warnings (like activity undertaken during Covid, with vaccine/medical disinformation). We welcome the steps taken as part of the Online Charter but would like to see them encoded into a more formal regulatory requirement with very clear targets and key performance indicators to track the impacts.

Effective oversight and a robust enforcement approach to reduce harm

- The nature of Fraud and Scams in the UK is ever changing, with well organised, funded and sophisticated crime gangs evolving their tactics and methods on an almost daily basis. This often results in significant harm in a relatively short amount of time. As a result, we strongly believe that Ofcom need to regularly review their risk profiles at least every six months, to ensure the regime remains nimble against evolving criminal methodologies.
- Ofcom must be willing and able to take a robust enforcement approach to drive improved consumer protection, to ensure platforms become safer environments. Enforcement action where identified failings have led to users experiencing harms should include fines that can be used to reimburse victims *of fraud* or develop *fraud prevention* technology to reduce harm across the eco-system.
- As a minimum large firms and multi risk online services should be required to have independent audits to help them drive consistent approaches and further understand best practices in place across multiple industries or ecosystems where consumer protection controls are required. Within the FS sector the second payments service regulation includes annual audits requirements, which has improved the implementation of the regulations.

---

[3] https://www.gov.uk/government/news/major-campaign-to-fight-fraud-launched#:~:text=Fraud%20accounts%20for%20around%2040,billion%20in%20England%20and%20Wales.

Volume 2: The causes and impacts of online harms?

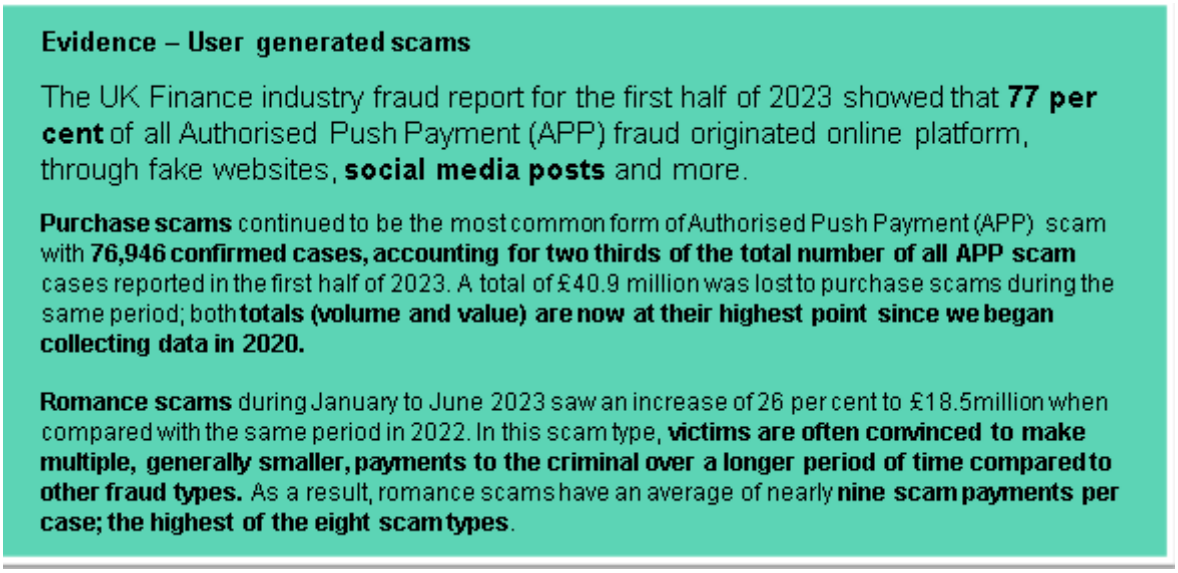
The FS sector primary areas of concern for volume 2 are as follows:

- Online Services that facilitate high levels of fraud due to user anonymity or pseudonymity such as social media, online marketplaces and dating sites need to have Identity and Verification (ID&V) requirements in place as a minimum. This would significantly limit a range of harms including impersonation, purchase and romance scams by making it easier to identify perpetrators of harm and, most importantly limiting their entry onto Services, through fake profiles, in the first place.
- Criminals use social engineering techniques to instil trust within their potential victims. Verified statuses and/or the impersonation of trusted brands or persons is a significant area where online services should be implementing stronger safeguards.
- Several of the Ofcom 'Risk Factors' do not have fraud and financial services offences captured as potential illegal harms; these are outlined later in this section.

Q) Do you have any comments on Ofcom's assessment of the causes and impacts of online harms? Do you think we have missed anything important in our analysis?

While the causes and impacts of online harms are important, the scale of the illegal harms committed against users, and the online services targeted prolifically by criminals should also factor into considerations. The volume of criminal attacks within this space is unprecedented with tens of thousands of victims facing attacks in a matter of weeks, and approximately 400 purchase scams occurring every day.

The Financial Services sector observes that **online services are being used as hunting grounds for potential victims**. Criminals take advantage and can attract potential victims in plain site, where pseudonymity and anonymity can be leveraged to commit fraud and scams. Criminals often face little to no limitations or controls when creating attractive or enticing profiles or posts to induce potential victims of fraud and scams. Purchase Scams, Romance Scams and mule recruitment are prevalent in the User-to-User environments, and occur where the users' posts, and profiles can often be seen by all/many users. Within some Fraud Modus Operandi (Investment, Romance, Advanced Fee, Bank/Police Impersonation), once a victim is engaged, they are subsequently convinced to leave the introductory environment and are moved off platform, where the criminals build a relationship and socially engineer the target victim with a ruse, convincing them to part with their money. We believe that preventative and proactive activity being undertaken by online services is fundamental to preventing harm to users and tackling the upstream origination issues that are impacting the FS sector.



4

Q) Do you have any views about our interpretation of the links between risk factors and different kinds of illegal harm? Please provide evidence to support your answer

The risk factors below do not mention fraud and financial service offenses; however, these illegal harms should be drawn into the risk profile examples. Criminals are pervasive and adapt to exploit any viable opportunity to defraud users, often existing within the grey areas of regulatory and legal instruments and guidance.

- **Risk factor: Livestreaming –** can be used in romance scams
- **Risk factor: Hyperlinking –** can be used in purchase scams and advanced fee scams to harvest credit/debit card credentials and of user information for social engineering at a later date. Bots are often used to generate image based and text-based posts where victims are led to believe they have the opportunity to receive a limited time discount or offer and so they must "purchase" / share their card details willingly, as soon as possible.
- **Risk factor: Discussion forums and chat rooms –** can be used to facilitate fraud and financial services offences such as Arranging which is another offence under POCA that relates to facilitating the transfer or disposal of assets that have been obtained through criminal activity. This includes providing advice or assistance to individuals who are looking to move or conceal assets. They are also used to facilitate private communities within investment scams where victims are convinced to make large investments, validated by the success of others within the groups (often a second criminal or a single criminal using multiple accounts to create an illusion of community).
- **Risk factor: User profiles -** can be used to facilitate fraud and financial services offences such as romance scams and investment scams, not all criminals stay anonymous.
- **Risk factor:** Anonymous user profiles or users without accounts can be used to facilitate fraud and financial services offences and arranging under POCA as this includes providing advice or assistance to individuals who are looking to move or conceal assets.
- **Risk factor: Commenting on content –** could be used to post split URLs/contact details for discounts etc, which can lead to harvest credit/debit card credentials and of user information to social engineering them later. This can also play a role in being the first point of contact a user

[4] https://www.ukfinance.org.uk/system/files/2023-10/Half%20year%20fraud%20update%202023.pdf

has with a fraudster, under the guise of mutual friends/interests, supplementing the anonymity and pseudonymity advantage as above.

Due of the way fraud originates across the ecosystem **we strongly recommend that the 'enhanced inputs' below is moved into 'core input' for the illegal harm of fraud.**

- 'Views of independent experts' which includes experts on industry trends, regulatory standards and the views of certain trade bodies or technical experts in relevant fields.

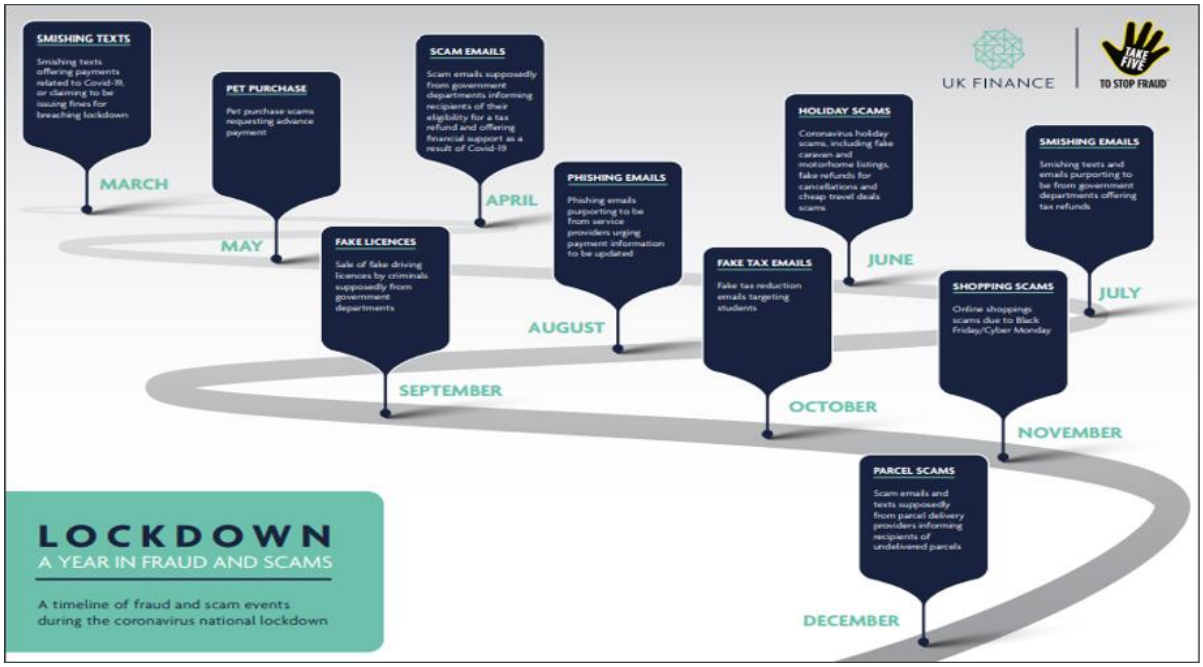
The FS sector uses a suite of tools including but not limited to device and, behavioural analytics, biometrics, reference data, transaction modelling systems. The online services should be proactively leveraging intelligence from the FS sector and vendors where appropriate, to be able to react to illegal content in real time. Also, within the scam's environment an initial scam approach can lead to multiple outcomes for example a Romance Scam approach can lead to several illegal activities (see Appendix 1 illegal harms). Finally, where an online service can identify users that have been exposed to illegal content related to fraud and financial service offenses (through trusted flaggers or user reporting), they should be required to share the information with the FS sector to safeguard the potential victim from fraud.

Volume 3: How should services assess the risk of online harms?

The FS Sector primary areas of concern for volume 3 are as follows:

- The nature of Fraud and Scams in the UK is ever changing, with criminals evolving their tactics and methods on an almost daily basis. This often results in significant harm in a relatively short amount of time. As a result, we strongly believe that Ofcom need to regularly review their risk profiles every six months, to ensure the regime remains nimble against evolving criminal methodologies.
- There is also a need for event driven risk assessment and risk profile reviews, whereby trusted flaggers or user reporting indicates a sudden high volume of a particular new type of scam or emergence of a new MO.
- There is a lack of reference to real-time data/intelligence sharing within the sectors to prevent repeat attacks and also data sharing across sectors to protect mutual users. See Annex A
- We believe there is a need for horizon scanning, scam/fraud migration assessment and the use of widely available information to proactively identify new criminal activities. Criminals are pervasive and adapt their techniques to circumvent controls at pace, the online services that are misused need to proactively review the evolving threat landscape.
- We recommend independent audits annually during the initial phase of the OSA. In particular where some online services hold a disproportionate level of risk. The culture within online services needs to change towards prioritising the protection of users. The OSA fundamentally challenges disregard for harmful outcomes, as such it is essential that those at the top of these organisations are receptive and adaptable to change.
- The measures should feature both qualitive and quantitative information as part of assessing the risks of online harms.

Having an illegal harm event or new adverse MO emerging as a trigger for risk assessment review will ensure online platforms are at the forefront of illegal harms prevention. Online scams can spread quickly and impact significant numbers, so it is important to have some form of event driven risk assessment requirement and review of Risk Profiles. Ofcom can set proportionality thresholds for this. The image below, illustrates how FS have identified the seasonal nature of fraud and scams and how MOs and events diversify throughout the year.



- Automated data/intelligence sharing within and across sectors impacted by fraud facilitated through the online services would drive a step change in tackling the criminals.
    - The trusted flagger option facilitates the sharing of typology, trends and statistical evidence. However, where online services are prolifically exploited by criminals, more real time (API based) data sharing will aid the mitigation of online harms. An example where this exists is between FS and law enforcement; both UK Finance and CIFAS share data with law enforcement at scale to support their link-analysis of crimes.
    - Trusted flaggers expansion for FS intelligence input is vital due to the significant amount of real time insight into frauds that are taking place on the online service platforms which can immediately alert online services to specific MOs or fraudulent adverts/tactics, so that they can take effective and timely action.
- There also needs to be horizon scanning and/or obtaining widely available information from trusted flaggers or counter fraud agencies, to proactively identify new criminal activities. For example, when the FCA verification rules were applied to the online advertisers there was an upward trend of account takeovers of genuine advertising accounts.
- Two-way information/data sharing is also required to stop the criminals leveraging the victim pools via online service providers. The criminals socially engineer victims to circumvent a number of bank controls to defraud their target victims; the online services will have high risk indicators that should be shared with the FS sector. Other areas where sharing information and intelligence back to the FS sector are recommended as follows:
    - Sharing live examples of adverts/user posts that have been taken down due to suspected fraudulent activity. Ensuring that any company names/data in the adverts/links are also shared. Data about how many clicks the advert/post received, and how long it was up before getting taken down, should also be shared so FS firms can understand the potential volume of victims in an attack.
    - Sharing sort codes / account numbers that may be ‘suspicious’ because they have been included in suspected communications between fraudsters and victims.
    - Sharing anything that could be mule recruitment etc. (see Appendix FS illegal harms) for the mule herder archetypes used to entice users.
- The large firms and multi risk online services should be required to have independent audits to help them drive consistent approaches and further understand best practices in place across multiple industries or ecosystems where consumer protection controls are required. Within the FS sector the second payments service regulation includes annual audits requirements, which has improved the implementation of the regulations.
- Fraud accounts for 40% of reported crime in England and Wales, and the majority of scams begin via online services. As such fraud should be in scope as a trigger for the proposed measures for automated content moderation and user access restrictions, as opposed to only applicable for CSAM or terrorism. As a trigger this could be combined with other data points to make informed and proportionate decisions.

We outline our specific concerns and recommendations below:

- For the Governance and Accountability, we welcome a nominated person for all online services, but note that the proposals focus on tracking, whereas this should include greater references to mitigation and measurement specifically. Proactive mitigation and the prevention of incidents recurring should be encouraged within the measure wording.
- The scope of Ofcoms’ expectations on monitoring needs to be made clearer in the measures. The measures, governance and accountability do not always appear achievable or measurable, it is unclear how an online service can write test procedures to evidence compliance.
- Services such as dating sites would only have to undertake having a nominated person which is disproportionate given a romance scam can lead to investment scams, money muling and sextortion.
- Also, the example of cars where the same images are used across different profiles would not be captured. Whereas proactive use of technology such as AI could help platforms detect

high risk activity for subsequent human review. In the Home Affairs Committee (HAC) evidence session held on the 7th of February Meta quoted that hate speech had been halved using this technology[5].

Measure summary (*source Ofcom illegal harms consultation package*)	Scope
Governance & Accountability	
Boards or overall governance bodies carry out an annual review and record how the service has assessed risk management activities in relation to illegal harms, and how developing risks are being monitored and managed	Large Services
A named person is accountable to the most senior governance body for compliance with illegal content safety duties, and reporting and complaints duties	All Services and All Risks Profiles
Written statements of responsibilities for senior members of staff who make decisions related to the management of online safety risks	Large Services **OR** Multi Risk
Internal monitoring and assurance function to independently assess the effectiveness of measures to mitigate and manage the risks of harm, reporting to a governance body or an audit committee.	Large Services **and** Multi Risk
Evidence of new kinds of illegal content on a service, or increases in particular kinds of illegal content, is tracked and reported to the most senior governance body	Large Services **and** Multi Risk
A Code of Conduct or principles provided to all staff that sets standards and expectations for employees around protecting users from risks of illegal harm	Large Services **and** Multi Risk
Staff involved in the design and operational management of a service are sufficiently trained in a service's approach to compliance	Large Services **and** Multi Risk

The content moderation measures for human moderation are too reactive, where the illegal guidance document outlines additional checks to determine if a piece of content is illegal. This is limited to the posts immediately prior and following the content being assessed, and the profile activity. This does not appear to include where the same content is reused. Also, criminals that evade detection often post enticing photos without text and explain the illegal activity via DMs to evade detection and takedown. Proactive engagement of human moderators based on customer complaints or trusted flaggers would determine the intent of the profile owner. Often there are posts reported where there is intelligence of criminal activity, however the online services automated systems respond to users with there has been no breach of their T's & C's' as the automated systems have not recognized the criminals' intentions. The reporting processes need to allow for adequate context as the criminals are circumventing the controls in place for example by leveraging images or posting URLs that appear innocuous.

[5] https://committees.parliament.uk/oralevidence/14233/pdf/

Content Moderation – Manual checks[6]	
Content moderation systems or processes are designed to take down illegal content swiftly	All Services and All Risks Profiles
Internal content moderation policies are set having regard to the findings of risk assessment and any evidence of emerging harms on the service	Large Services **OR** Multi Risk
Performance targets are set for content moderation functions and services measure whether they are achieving them	Large Services **OR** Multi Risk
Content moderation teams are resourced to meet performance targets and can ordinarily meet increases in demand for content moderation caused by external events	Large Services **OR** Multi Risk
Staff working in content moderation must receive training and materials to enable them to identify and take down illegal content	Large Services **OR** Multi Risk

For the Automated Content moderation, the limited application to CSAM offenses lacks proportionality given the criminals are using gangs, bots and AI to post contents across online services. These measures should have risk factors that relate to fraudulent contents to increase the efficacy against wider illegal harms:

- There should also be thresholds which trigger proactive technology and automated content moderation, on a risk-based approach, to determine underlying illegal harms.
- There is a need for far more proactive technology to identify URL hosting malicious content and or harvesting data in User Generated Contents – as was seen in the Wilko scam incident (see Appendix 1 illegal harm examples). Also, keyword searches that are strongly associated with offenses would miss the unusual but legal terms such as 'clean title' which would be a 'red flag'.
- There are vendors that specialise in removing URLs that have nefarious contents such as phishing, malware for the FS sector, these have been proven to significantly improve an online services fraud and scam detection rates in a trial. There are also vendors and services such as https://www.scamadviser.com/ which assess URLs to help identify if a website is safe, legit or a scam. Automated URL checking based on set triggers would be a proportionate control online services can implement.

In the Home affairs committee evidence session on fraud Meta noted the use of AI had halved Hate Speech, this technology could be trained to determine illegal or suspicious activity, this could be leveraged to alert for human review if there is a false positive concern[7].

As such we believe the Keyword search measure needs to be expanded to capture images and URL checking, as well as taking in a trusted flagger or risk profile examples of red flags to search for reoccurring content across an online service. It is not proportionate to alert each individual post an online service to mitigate organised criminal gangs posting high volumes of repeat content.

[6] source Ofcom illegal harms consultation package

[7] https://committees.parliament.uk/oralevidence/14233/pdf/

Automated Content Moderation Automated[8]	
An automated technique known as 'hash matching' is used to detect image-based Child Sexual Abuse Material (CSAM)	Apply to the specific risk of CSAM
Automated tools detect URLs which have been previously identified as hosting CSAM or which include a domain identified as dedicated to CSAM	
Keyword search is used to detect content containing keywords strongly associated with offences concerning articles for use in frauds (such as the sale of stolen credentials)	Applies to fraud

The reporting and complaints processes are only as effective as the actionable information gathered. Current experience of reporting is that there are generic categories with no further options to provide context, and this frequently results in automated response that there is nothing against the guidelines of the online services. This is due to of a lack of granular information being captured within the reporting processes. This is precisely how criminals are able to exploit online services unfettered. Context is vital to designing out the criminal activity, and trusted flaggers could assist online services determining the balance of information required to effectively mitigate criminals that prolifically exploit them.

Reporting and complaints Measures[9]
Complaints processes enable UK users and affected persons to make each type of relevant complaint in a way which will secure that appropriate action is taken
Complaints system & processes are easy to find, easy to access and easy to use
Appropriate action: indicative timeframes for considering complaints should be sent to complainants
Appropriate action for complaints: illegal content complaints should be handled in accordance with our proposed content moderation recommendations
Appropriate action for complaints: performance targets for determining appeals should be set and services resourced to give effect to them
Appropriate action for complaints: appeals are determined promptly
Appropriate action for complaints: upheld appeals should lead to the complainant being restored to their original position
Appropriate action for complaints: for proactive technology complaints, the service should inform the complainant of their rights
Appropriate action for complaints: other complaints should be triaged and passed to the appropriate function or team internally, with view to protecting users from harm

UK Finance stress the need for public transparency around what the different types of verification statuses actually mean. As evidenced previously, these schemes are abused by criminals to impersonate and enhance their trustworthiness, which are pillars within social engineering attacks. Ofcom should place policy restrictions under which these statuses cannot be acquired/maintained, such as being on the FCA watch list for promoting fake investments.

[8] source Ofcom illegal harms consultation package

[9] *source Ofcom illegal harms consultation package*

Enhanced User Control[10]	Scope
There are clear internal policies for operating notable user verification and paid-for user verification schemes and improved public transparency for users about what verified status means in practice (e.g. Blue Ticks)	All Services and All Risks Profiles

There are numerous examples where fraud, POCA and FSMA should be within scope of the user access measure, to aid proportionality and reduce false positives; these examples could be 'red flags' that are triggers for the User Access measure to be utilised.

These examples include but are not limited to:

- Prolific mule or mule herder
- Linked to prolific fraud indicators and complaints from consumers/trusted flaggers
- On the FCA watch list

User Access	Scope
Accounts should be removed if there are reasonable grounds to infer they are run by or on behalf of a terrorist group or organisation proscribed by the UK Government	This only applies to the specific risk of terrorism

Q) Do you agree with the types of services that we propose the governance and accountability measures should apply to?

To carry out certain regulated activities in financial services, entities need to have a licence, regardless of their size. This model uses a risk-based approach across channels, products, locations etc. to determine the appropriate mitigation of risk. For the OSA illegal harms a risk-based approach or outcome-based approach would at least capture where the size of firm is smaller, such as an online dating site. Currently a platform with 9% of the UK population as users could be subject to solely having a nominated person as part of their governance regime, where they have one risk type. This does not seem a proportionate approach to mitigating the illegal harm of fraud.

It would also be important to ensure the size of firm cannot be manipulated to avoid considerable new obligations, if it were classified as a 'large' online service. A platform could split entities to avoid obligations under the OSA in the UK.

Ofcom could consider using the case reports (confirmed financial losses) on Authorised Push Payment (APP) Scam, from Financial Service Institutions, to assess a Services adherence to the act. To go one step further, this comprehensive data could be used form the equivalent of a user complaint, giving Ofcom valuable insight to help assess whether individual companies are doing enough to protect their users from illegal harms. A breakdown of confirmed scams cases reported per online service will be shared with Ofcom [redacted] to aid further consideration on the application of the code of practice and the measures.

Q) Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to requiring services to have measures to mitigate and manage illegal content risks audited by an independent third-party?

The current lack of independent audit is a missed opportunity to follow best practice. In the FS sector the second payments directive regulation for strong customer authentication includes annual

[10] *source Ofcom illegal harms consultation package*

audits requirements (which helped services improved the implementation of the regulations.) This led to best practices, more consistency and therefore more robust interpretations of the SCA regulations.

Q) Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to tie remuneration for senior managers to positive online safety outcomes?

It has taken multiple years to investigate complex cases when the equivalent FS regime was introduced. The FCA are well placed to provide Ofcom with best practices, influencing strategies and enforcement including the usage of fines. See below for the volume of cases carried out since the FS SMCR regime came into place.

Example - FCA SMCR enforcement

In total, the FCA has over 370 individuals under investigation for a range of potential breaches.

		2016	2017	2018	2019	2020	Total
Financial Penalty; Public Censure	Senior Manager	0	0	1	0	0	1
	Non-SMF individuals*	0	0	0	0	0	0
Prohibition	Senior Manager	0	0	0	1*	0	1
	Non-SMF individuals	0	0	0	0	0	0
Undertaking**	Senior Manager	0	0	1	0	0	1
	Non-SMF individuals	0	0	1	0	0	1
No further formal enforcement action***	Senior Manager	0	0	3	1	11	15
	Non-SMF individuals	0	0	4	9	15	28
Total		0	0	10	11	26	47

11

[11] https://www.fca.org.uk/freedom-information/information-senior-managers-certification-regime-fines-april-2022

Volume 4: How to mitigate the risk of illegal harms –the illegal content Codes of Practice

The FS primary areas of concern for volume 4 are as follows:

We do not think the approaches to mitigate the risks of illegal harms are robust enough to keep pace with the criminals that are exploiting the online services for fraud, FSMA and POCA. Our recommendations are below:

- Peer-to-peer marketplaces without adequate controls are driving a significant level of harm to UK consumers. Users are more likely to fall victim to purchase scams than any other type of scam, and these services are where the majority of purchase fraud happens - with unvetted and unregulated sellers accessing UK consumers en masse. TSB estimates that 73% of all purchase scams they see are driven from Facebook Marketplaces. These platforms need to introduce tailored controls, strengthening the voluntary commitments in the Online Fraud Charter, specifically:
    - Verification of sellers: anonymity makes it easier for fraudsters to list fake items, as the barrier to entry (i.e., simply opening an account) is so low. Requiring verification raises the bar for criminal to gain access to these online marketplaces.
    - Integrating with secure payment service providers: large firms, e.g., Facebook, must be required to integrate with secure payment service providers to offer their users a safe way to pay for goods and services online.[12]
    - • Greater use of proactive technology is required to strengthen effectiveness (including but not limited to AI, machine learning, analysis of linguistic and stylistic cues, image crawlers), as well as real time information and intelligence sharing both within and across sectors.
- Automated content moderation and user access measures in the proposal have been scoped too narrowly to CSAM or terrorism only – other illegal harms (including fraud) should be brought into scope of these measures too.
- We do not agree the size of large firms being defined as 10% of the UK population is proportionate, as we believe this would completely exclude online services that are leveraged by criminals to exert heinous romance scams for example. A risk-based approach would be better, or additional risk factors/profile that captures online dating services with greater measures requirements.
- An effective step to prevent criminals from operating on services tackle the Fraud related harm encountered on online services would be to require firms that allow anonymity or pseudonymity (such as social media, online marketplaces and dating sites) to have to undertake Identity and Verification (ID&V) amongst their controls to prevent or at worst hinder bad actors' ability to enter the ecosystem.
- Dating Sites need to be captured under more of the larger firm measures than is currently proposed, given the significant role that they play in facilitating romance scams. In this scam type, victims are often convinced to make multiple, generally smaller, payments to the criminal over a longer period of time compared to other fraud types. As a result, romance scams have an average of nearly nine scam payments per case: the highest of the eight scam types. Dating sites recommendations include performing ID&V on all of their users; sharing real time data with FS sector regarding suspicious activity and being required to close the accounts of anyone purported to be setting up fake accounts or accounts being used by fraudsters.

[12] See the case study for improve marketplace controls.

Q) Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views

The proposals made by Ofcom are not commensurate to the criminals' adaptability and aggression; there are a number of recommendations we have made that would have greater impact mitigating criminal activity. We have also commissioned independent consumer research on what more enabling sectors of fraud and scams can do to mitigate the consumer harm, which is outlined in detail within appendix 2 Thinks research.

Q) Do you have any comments on our overarching approach to developing our illegal content Codes of Practice?

Whilst the approach to developing the illegal content codes of practice is logical, the regime is vast and the consultation package immense. It would be beneficial to provide additional shorter guides on the illegal harms for different online service types or harm types, to aid stronger implementation across the sector. We believe a series of sanitised visual examples for different illegal content types would bring the spirit/intention of the regulations to life for smaller players[13].

Q) Do you agree that in general we should apply the most onerous measures in our Codes only to services which are large and/or medium or high risk?

The FS sector operates under a Risk Base Approach (RBA) regime, this is irrespective of the size of firm. Controls are tightened dependent on the risk profiles identified by the FS firm, there is threat landscape monitoring to understand the migration of criminals and fraud types, horizon scanning to increase or develop controls. The measures within the codes of practice should be applied on a risk-based approach to illegal harms that are prevalent or viable on the online service. Some large online services have disproportionately low fraud/scams rates, this may be due to the measures being in place already and therefore the measures are not onerous as below:

Case Study - Online Marketplace control improvements

- 150 million+ listings and messages shared on the platform in 2022.
- By leveraging updated safety infrastructure throughout 2022 they removed 6.32% of listings from the platform – up from 3.99% the previous year.
- In 2022 99% of the listings were without issue.
- Just 0.02% of the listings were reported for fraudulent / harmful behaviour – an all-time low.
- Improved safety tools automatically blocked 17% more listings and 28% more bad accounts.
- Suspicious listings / messages reported by users were investigate by the moderation team within 4 hrs.
- 24/7 support was provided to the community via the Trust and Safety team.
- Swift action was taken against listings that fell outside the rigorous Policies and Code of Conduct to keep the community safe.
- Strengthened ability to stop bad activity, an additional 746,380 listings were automatically deleted from our platform in 2023, a 17% YoY increase. The listings were removed because the user was blocked, or the content fell outside our posting policies.

Total listings removed by category

Category	2021	2022	2023	YoY % Change
Cars, Vans & Motorbikes	183,635	141,589	116,441	-18%
Community	13,939	15,859	12,509	-21%
Flats & Houses	52,390	49,149	51,090	+4%
For Sale	679,948	1,357,626	933,134	-31%
Pets	128,694	102,549	55,484	-46%
Services	15,722	14,494	8,634	-40%
TOTAL	1,075,339	1,681,266	1,177,292	-30%

Removal reasons

Reason	2021	2022	2023	YoY % Change
Spam	82,610	418,369	156,357	-63%
Following user reports	139,584	120,107	88,484	-26%
Fraud & Suspected Fraud	41,286	62,156	42,725	-31%
Prohibited	86,049	53,396	44,195	-17%
Counterfeit	7,658	17,721	7,066	-60%

[13] See FCA financial promotions guidance document in https://www.fca.org.uk/publication/finalised-guidance/fg15-04.pdf which clearly demonstrates what is compliant and what is not compliant.

Q) Do you agree with our definition of large services?

The definition of large services where it has an average user base greater than 7 million per month in the UK, approximately equivalent to 10% of the UK population would mean that most if not all dating services would not be captured. Yet it is clear a romance scam can lead to many illegal outcomes and most dating online services would not be in scope of the measures. For the FS sector financial crime and risk frameworks are a baseline requirement to acquire a license from the FCA, we believe the current definition of a large service is not a proportionate approach to managing the harms.

Q) Do you agree with our definition of multi-risk services?

Fraud is a predicate offense to money laundering, as such a money mule can span from fraud to POCA offenses, this could therefore be classified as triggering the two illegal harms. Likewise illegal/ fraudulent half price goods sales and purchase scams would likely trigger multiple illegal harms. We are concerned that small/medium firms may not understand where they are in fact multi risk services. We would recommend that Ofcom produce examples or guidance which clearly articulates common areas of ambiguity for those online services that are exposed to the risk of these illegal harms.

Q) Do you have any comments on the draft Codes of Practice themselves?

The codes of practice and regulatory products reference many options of resources to enable the successful takedown or removal of content, but the more specific detailed guidance is hidden within numerous annexes. The calls to action should be made more prominent as these are the preventative measures that could be undertaken. We do not believe the current materials are accessible to small, micro businesses which may be multi risk.

Q) Do you have any comments on the cost's assumptions set out in Annex 14, which we used for calculating the costs of various measures

The FS sector spends significant amounts of funds to protect users from economic crime and as such we believe the user harm should play a significant factor relative to cost to the business. See below for some analysis of the banking sector

Example of costs - Banking sector spends more on IT than others sectors

- Financial service firms have to fulfil exacting regulatory requirements which translate into IT costs that do not contribute to the firms' earnings.
- Financial crime compliance spend for 2022 equivalent to three quarters of UK defence spend. Regulation remains the biggest perceived external compliance cost driver – more so than financial crime itself.
- Main internal compliance cost drivers are increased automation, data, tools and new technologies, as well as growing financial crime compliance volumes.
- Biggest costs commitments over the next three years predicted to be transaction monitoring , KYC/IDV and fraud checks at onboarding.

[14]

[14] https://risk.lexisnexis.co.uk/insights-resources/white-paper/true-costs-of-compliance; https://www.finextra.com/finextra-downloads/featuredocs/high_price_of_it.pdf

Volume 5: How to judge whether content is illegal or not?

- The judgment guidelines are too heavily weighted on internal information points within the online services systems; this is a blinkered view as some of the illegal activity indicators occur outside of platform, such as fraud. Data sharing is critical for effective mitigation on some illegal harms.
- The three groups of reasonably available information on a platform are not ambitious relative to the criminal's aggression, also the fact that group one is a prerequisite will leave gaps. The examples given are either too narrow, or focus on written content rather than images, calls and live streaming services.
- Financial Services Trade Bodies and Specified Anti-Fraud Organisations with an intelligence function (such as UK Finance, Cyber Defence Alliance and CIFAS) ***receive invaluable data from its membership on key threats***. This intelligence resource should see them included as Trusted flaggers to help all participants in the fraud ecosystem keep pace with the criminal actors and their adaptable nature.
    - We also recommend, as a minimum the banks and building societies with over 7mn users should be independently eligible for trusted flagger status, as these entities mirror the OSA 'large' service definition, **have specialised knowledge and intelligence to share that comes from encountering tens of thousands of victims firsthand**.
    - One member has suggested this should also be extended to Trading Standards be eligible also
- Trusted flaggers and counter fraud groups should be classed as reasonably available information sources.

Q) Do you agree with our proposals, including the detail of the drafting? What are the underlying arguments and evidence that inform your view.

We are very concerned that criminals use bots and AI to generate high volumes of contents, and the automated content detections are primarily being limited to CSAM related harms. As such the automated content detection technologies being proposing in the codes fall short of effective illegal harm mitigation for fraud, FSMA and POCA. We would recommend wider technologies to mitigate FS harms including but not limited to:

a) Repeat imagery related to fraudulent posts should be a risk indicator, this is apparent where the same imagery is used to create dozens of replica profiles/posts.
b) The financial services sector has used many suppliers to eliminate impersonation and false representation URLs by criminals. These proactive technologies could scan UGCs in high-risk environments, to prevent victims being exposed to criminal actors.

Whilst the keyword search will mitigate the criminal activity that is commonplace, the lack of imagery checks would miss the examples found for car purchase scams. Also, this will only combat specific sub offences, content promoting articles for use in fraud, rather than the priority offense of fraud as a whole as described in Schedule 7 of the Act. We recommend Ofcom consider expanding the requirement for online services, specifically large firms with a high risk of fraud, to require that they develop automated content moderation controls designed to proactively identify that high-risk content.

Bots posting content occurs routinely with online services; there has been examples where registration for a free event has triggered a bot to make a post offering a 20% discount if the reader follows a special link. This is used to harvest payment information from victims that fall for the discount promotion, this criminal activity would be more successful where an entrance fee is

required. The current reporting mechanisms do not provide sufficient flexibility to explain the sequence of events, as such reports against this type of contents do not appear to breach the online services terms and conditions.

Dating sites often have features such as chat and video functionality, however there are no online harms measures related to these features, nor a clear hand over to Ofcom's electronic communications regime. Deep fakes are proactively being used by criminals in romance scams, these online services could provide a safer environment to interact with each other by providing deep fake detection as part of their service offering to keep their customers safe.

Some high priority user generated scams statistics are outlined below, whilst automation is good with detecting activity at scale, human reviewers are good at reviewing for local context which is a critical component in tackling these types of scams. The distinction in capability between automated and manual moderation routinely leads to problems within the online marketplace, where words associated with criminal activity are often not used – as such the keyword search would miss some scams entirely. In this instance, image detection combined with user and or trusted flagger reports would be far more effective to removing and mitigating illegal contents from an online marketplace. And proactive technology detecting the recurring image posts would be preventative. as the criminals are using automation, the OSA measures Ofcom propose should be commensurate to the techniques/tools criminals use as a minimum.

Online Marketplace

UK Finance half year fraud stats showed that 77,000 cases of purchase scams were recorded by UK

banks in the first six months of 2023 alone – this is over 400 a day. Incidents are rising fast - the first half of 2023 saw a 43%3 increase in purchase scams, compared to the same period in 2022.

Bank A

- **Data covering January 2023 – September 2023, and January 2022 – September 2022, shows:**
- **440 customers have fallen victim to Facebook car scams this year, up 87% from last year. Customers in every age group reported an increase in car scams.**
- **This year nearly half a million (£479,964) has been reported as lost to Facebook car scams, up 93% from last year.**
- **People aged 18-25 are the most likely to fall victim to Facebook car scams accounting for a fifth (20%) of all cases in 2023. Over 60s are second most likely to fall victim accounting for 14% of all cases, while 41–45-year-olds are third accounting for 12.5% of all cases.**
- **Over 60s lost the largest sums of money to scammers, with an average claim made for £1,564 so far this year, up from £748.06 last year. 56–60-year-olds reported the second largest losses with an average claim for £1,528, while 26–30-year-olds were third with an average claim of £1,263.**

Bank B

- **Scam cases originating on Facebook Marketplace account for 77 percent of all purchase scams at Bank B. Instagram follows, with nine percent then Twitter (4%), Snapchat (3.5%) and eBay (2%). Bank B highlights that eBay's payment platform and customer verification has led to a miniscule fraud rate, compared to Facebook Marketplace.**
- **Online marketplace controls - A banks research determined, that for 100 sellers they engaged with, 30% of sellers were fake in addition to this, sellers were targeted with a range of scams via chat to persuade them to part with cash. The platform could easily set up test seller accounts to detect criminal actors on the face book marketplace.**

Bank C

The impact of payment an integration is clear in our data: platforms with purchase protections in place drive fewer fraud losses for our customers. In the last 6 months of 2023, customers reported just under 4,000 cases of purchase fraud originating from Facebook. In comparison, Vinted, a peer-to-peer marketplace that integrates with a secure payment service to allow their users to pay for goods, drove just 670 cases. This cost neutral solution would help tackle the most prevalent APP fraud type in the UK.

Online marketplace mitigating controls for sellers

One bank researched and determined where they engaged with 100 sellers, 30% of sellers were Fake. In addition to this, sellers were targeted with a range of scams trying to trick sellers to click rogue links (in order to capture their log in credentials) or to potentially dupe them into releasing payment credentials under the ruse of fake insurance. It would be rudimentary for an online service to test up a series test account and find criminal approaches within their platform, to proactively mitigate these attacks.

Trends associated to fraud and scams

Understanding where the frauds start can aid benchmarking, if a dating platform is double the size but has half the victims there may be best practice available to improve wider industry controls. The attack levels as well as the victim levels are important factors to understanding effective controls

Romance Scams[15]

Case Studies (summary of three banks and UK Finance)

UK Finance data shows romance scam fraud increased last year with £31.3 million worth of romance scams reported in 2022, up from £30.9 million in 2021, and up from £17.8 million in 2020.over time.

- **Romance scams rose by 22% last year, with an average £6,937 stolen.**
- **People aged between 55 and 64 most likely to fall victim. However, it is those aged between 65 and 74 who lose the most money, giving romance scammers an average £13,123, the highest amount of any age group.**
- **83% who fall victim to romance fraud do so due to scammers' clever choice of words The phrases to watch out for include: (I've fallen for u, My £££'s frozen, I'll pay u back, I can't video call, We're so alike, Trust me, Only u can help, We'll be married,U know me, Soz, I'm abroad.)**
- **Almost a third (31%) of Brits have been targeted by a romance scammer.**
- **Brits who have fallen for romance scams say they lost £2,300 on average**
- **More than four in five Brits (83%) who fell victim to a romance scam said it was because of the clever language used by the criminals, the way they were spoken to, or the intimate conversations they had with the scammer.**
- **Facebook - where fake profiles led to over a third (35%) of all fraud cases.**
- **This is followed by almost a quarter (24%) on Tinder, over a fifth (21%) on Plenty of Fish and almost one in 10 (9%) from Match.com.**
- **The following platforms services all account for three percent of cases in which the platform was recorded:** Olderdating.com; Bumble and Instagram.

There is rich intelligence and information from sectors such as FS which is critical to informing context and effective judgements for online safety. The activity considered within the three groups below, will not help online services stay ahead of the criminals. We believe the proposals could go further to create a more impactful regime as below:

[15] https://www.express.co.uk/finance/personalfinance/1863137/lloyds-bank-romance-scam-warning; https://www.santander.co.uk/about-santander/media-centre/press-releases/santander-partners-with-dating-expert-to-warn-of-the; https://www.tsb.co.uk/news-releases/tsb-reveals-alarming-details-of-romance-fraud/

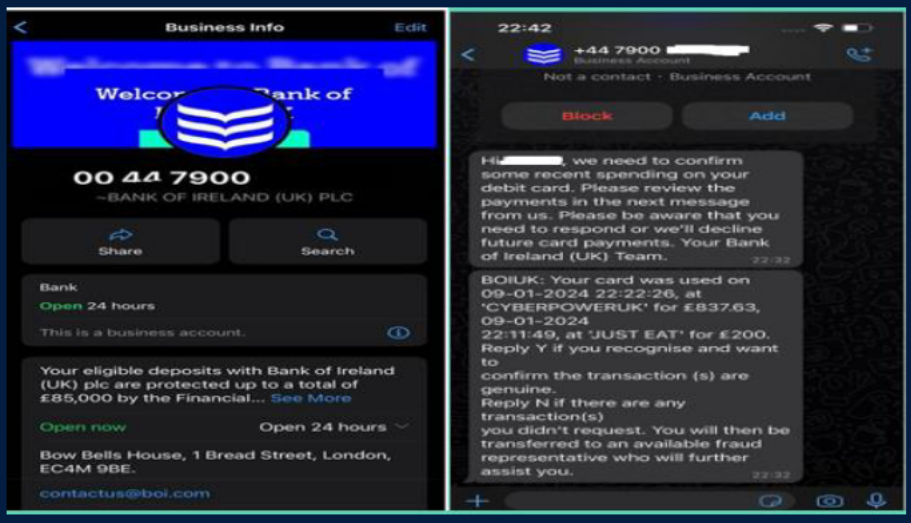
- **Group 1: Disguised account information or activity**
  - This group would always have to be a factor, and will cover a lot of circumstances, but it would not necessarily cover all. For example, a bogus company selling dubious goods/harvesting data/selling investments may not be classed as masking their identity.
- **Group 2: Requests, invitations, or inducements to invest, send money, send identification documents, or send financial information**
  - Mule herders often use lifestyle accounts to avoid contravening the terms and conditions; as such they would not openly ask for specific information in their User generated content, they often use lifestyle accounts with no words.
- **Group 3: Account and content characteristics commonly associated with fraudulent behaviour**
  - This does not capture any examples of images with or without words.

Impersonation of Financial Services

FCA has a list of authorised companies, it would be viable for online services to expand the FCA verification process for advertisers, to the User Generated contents environment to protect the FS from impersonation. Whilst the unique FRN and or specified person on FCA registers as proof of legitimacy is referenced. This would likely be used to validate after the content being posted impersonating that firm. The illegal content has to be spotted, usually after someone has fallen victim to then be reacted to, this is not a preventative approach. Proactively onboarding the FS companies would prevent impersonation of legal regulated entities by allowing them to own their footprint. Below is an example where a bank has been impersonated, and it has taken over 7 days to remove the criminal's fake account as below:

Case Study – Brand impersonation[16]

A WhatsApp Business profile impersonating Bank of Ireland UK was observed around January 10th, 2024. The profile was sending messages designed to look like card fraud alerts. The WhatsApp profile is shown below along with a screenshot of fraudulent messages received by a customer. The bank reported the profile to the online service through the in-app reporting mechanism but did not get any acknowledgement or reply. The profile remained live until the bank reported the profile URL as fraudulent through a takedown service provider.



[16] Source: UK Finance Member

The onboarding process would need to capture FRN and persons contact details in this scenario, otherwise the criminal can simply quote what is published on the FCA authorised list. Additional checks should be performed in these types of scenarios, where the criminals are impersonating, We would also recommend that the online services also using the FCA watchlists to determine if there is a clone or restricted FS firm in operation

Q) Do you consider the guidance to be sufficiently accessible, particularly for services with limited access to legal expertise?

We do not consider the guidance to be sufficiently accessible, in particular to small multi risk firms; there are a plethora of products that form the guidance (risk assessment, Illegal contents judgement guidance, risk profiles and codes of practice and the accompanying annexes).

To make these more accessible for smaller firms we believe there should be short guides based on functionality e.g., chat services, social media, online marketplaces and dating sites. Alternatively, the multi risk or specific risk online services should have one or two pages to demonstrate the linkages between the relevant sections of the codes of practice. All of these guides should be kept to a few pages, to provide clear overviews of duties and obligations. This would demystify the existing set of extensive and robust documents that could then been read in better context.

Alternatively, video content explaining from the functionality lens, as some smaller or micro business online services will be more familiar with the services, they offer than the particular illegal harms that may be present.

Q) What do you think of our assessment of what information is reasonably available and relevant to illegal content judgements?

We believe the reasonably available information sources are too limited, the examples of **reasonably available information** (content information, complaint information, user profile information, user profile activity and published information) fall short of those used within the financial service sector to stay ahead of criminals. Online services also need to consider information published by a "competent authority"; this could also include relevant competent sources. For example, the HMT National Risk Assessments for ML and TF & Proliferation Financing, NCSC, NCA, UK Finance annual fraud report, CIFAS Fraudscape, FATF, RUSI and FCA alerts emails which outline all warnings issued in the past 72hrs.

Trusted flaggers and counter fraud organisations should also be deemed reasonably available, as they would proactively engage and provide information to mitigate fraud and scams.

Volume 6: Information gathering and enforcement powers and approach to supervision

The FS primary areas of concern for volume are as follows:

- The nature of Fraud and Scams in the UK is ever changing, with criminals evolving their tactics and methods on an almost daily basis. This often results in significant harm in a relatively short amount of time. As a result, we strongly believe that Ofcom need to regularly review their risk profiles every six months, to ensure the regime remains nimble against evolving criminal methodologies.
- Ofcom must be willing and able to take a robust enforcement approach to drive improved consumer protection, to ensure platforms become safer environments. Enforcement action where identified failings have led to users experiencing harms should include fines that can be used to reimburse victims of fraud or develop fraud prevention technology to reduce harm across the eco-system.
- As a minimum large firms and multi risk online services should be required to have independent audits to help them drive consistent approaches and further understand best practices in place across multiple industries or ecosystems where consumer protection controls are required. Within the FS sector the second payments service regulation includes annual audits requirements, which has improved the implementation of the regulations.

Q) Do you have any comments on our proposed approach to information gathering powers under the Act?

As per the above, we believe leveraging the information gathering powers in the right way can drive good behaviours.

Ofcom enforcement should be aimed at really deterring online services to ensure that there is an incentive for services to take steps to prevent this content appearing. This could be achieved by publicly reporting enforcement action and outcome, similar to the FS and FOS roles, fines being imposed should also include the reimbursement of the victims, in particular where an online platform has made a profit in relation to the scam being carried out. Whilst it is promising to see that Ofcom anticipate full compliance within 6 months, we would urge that any case-by-case discretion has a cap on extensions to this. Not least as we have routinely observed, the last service to upgrade controls is often prolifically targeted by criminals.

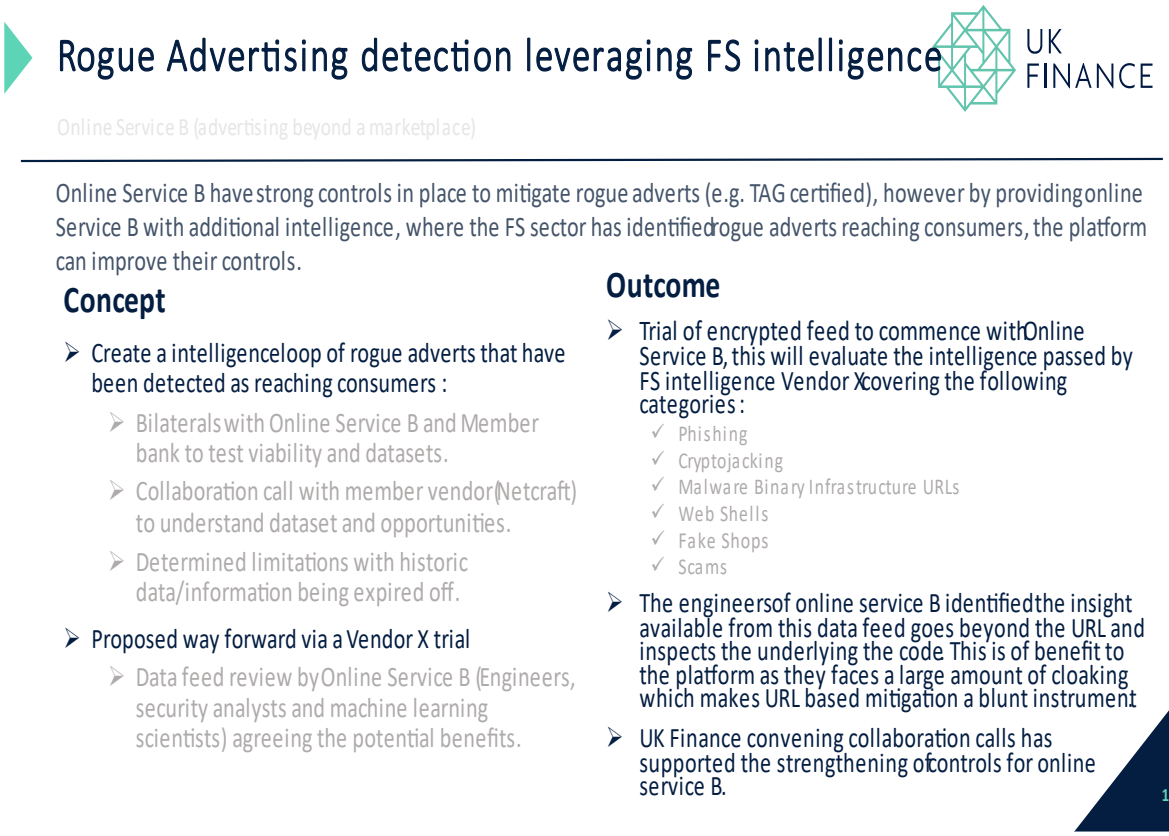
If you have any questions relating to this response, please contact Dianne Doodnath

Principal Remote Payment Channels Dianne.Doodnath@ukfinance.org.uk

Annex A – Online service deploying FS native vendors within the existing processes of online services

Platform feedback

- UK Finance members directly sees the “Voice of the Customer”, and can tell us what is actually getting through
- UK Finance members use vendors to mitigate scam contents, e.g., URLs/Adverts, so a platform using a feed from these FS specialist vendors we get direct information of scams that are successfully operating in the public domain, which we can feed to our detectors and block them at source, before they reach consumers
- Our testing projects determined an FS specialised vendor feed will give an online service a significant uplift in detections, primarily in the social-engineering space.
- These include scam yads (clickbait) as well as advertising–so improves quality as well as safety



The outcome of the **collaboration resulted in a 30% uplift in rogue advert detection/blocking rates**.