Ofcom Call for Evidence on the Online Safety Bill: UK Finance Response

Date: 13 September 2022

Answers to questions

1. Please provide a description introducing your organisation, service or interest in Online Safety.

5. Fraud poses a major threat to the UK public. According to UK Finance's 2022 Annual Fraud Report, members reported 195,996 incidents of Authorised Push Payment (APP) scams in 2021 with gross losses of £583.2 million.[1] Action Fraud's Annual Assessment of Fraud Crime Trends in 2020 and 2021 outlined social media services and encrypted messaging services as the key enabler of all frauds, and advertising via search engine optimisation as some of the key threats resulting in fraud and scams.[2] Furthermore, online shopping and auction frauds was seen as the second highest fraud type to be reported by victims.

6. Criminals use advanced social engineering techniques, and target users of online platforms by presenting them with malicious content designed to convince them to send money or divulge key personal or financial information. As the victim approaches and social engineering occur outside the Financial Services (FS) sector it is imperative that the online platforms actively intervene to protect consumers by preventing such malicious content and educate consumers.

7. Following the government's welcomed inclusion of user-generated fraud within 'the Bill', online companies will, for the first time, take responsibility for tackling fraudulent user-generated content on their platforms. As part of the imposed obligations as per 'the Bill', search engines, social media platforms and online marketplaces must implement adequate controls to prevent, detect and respond so as to mitigate the presence of fraudulent and scam activity across their platform, including the inhibition of migrations to other online platforms following control enhancements. A clear distinction needs to be drawn between these three segments to account for variances in harms identified and individual operating models.

8. Furthermore, the principle of "same activity, same risk, same regulation" need apply. Adopting this approach can help standardise the regulatory approach regardless of the nature of the provider and help to ensure consistent consumer protections.

   Please see below some distinct examples to demonstrate the range of harms based on platform type, some platforms may have multiple functions/types.

[1] UK Finance, *Fraud the Facts 2022. https://www.ukfinance.org.uk/policy-and-guidance/reports-and*https://www.ukfinance.org.uk/policy-and-guidance/reports-and-publications/annual-fraud-report-2022*publications/annual-fraud-report-2022*

[2] https://www.actionfraud.police.uk/fraud-stats

- **Social media** - can be used by bad actors purporting to be someone they are not or hiding behind usernames or fake identities so they cannot be tracked and disappear without trace. Examples of this would be money mule recruiters, fake celebrity endorsements or romance scammers.

- **Search engines** - can be abused when bad actors pay for advertisements to appear at the top of a search or use fake reviews so as to engender confidence in the integrity of their product or service. Examples would be fake / fraudulent investments.

- **Online marketplaces (OMPs)** - can be abused by sellers of fake, non-existence, and fraudulent goods. Where these occur on the online marketplace, the OMP must shoulder the burden of putting the customer right.

9. Additionally, the agnostic and global nature of fraud requires a collaborative and coordinated approach by all sectors to combat online generated fraud and scams. Responsibility for scam prevention should be equally shared by all parties for mitigation to be truly effective in protecting victims.

2. Can you provide any evidence relating to the presence or quantity of illegal content on user-to-user and search services?

10. The presence of fraud and scam activity transcends a multitude of platforms, covering a variety of different scam types as outlined below, the anatomy and effective mitigation of which need to be commonly understood. Losses due to authorised push payment scams were £583.2 million in 2021 [2], with over two thirds determined to have initially begun online via social media posts, online marketplaces, and online advertising. UK Finance has developed our secure industry case management system to identify and detail the source of each confirmed scam. Once a full quarter of analysis is available in Q4 2022, we will share this with Ofcom to demonstrate the trends and run rates per platform for the 8 Authorised Push Payment (APP) scam types that UK Finance tracks on behalf of the FS industry.

11. Criminals are increasingly targeting vulnerabilities that are outside the financial sector’s perimeter. Major online platforms – search engines, social media and online marketplaces - provide a gateway for fraudsters to identify and connect with their victims with fraudulent activity then occurring off platform. Online platforms must provide a greater barrier to fraud and not be used as a conduit to facilitate criminals' contacting potential victims. **Investment scams:**

- Members previously reported an increase in fraudulent investments advertised on search engines. Some platforms took early steps to proactively mitigate this by increasing due diligence to ensure adequate authorisation for financial promotions had been acquired by the Financial Conduct Authority (FCA). Following agreement by Online Fraud Steering Group (OFSG) members, Meta, Microsoft and Twitter have since committed to requiring UK regulated financial services to be authorised by the

[2] UK Finance, *Fraud the Facts 2022.* *https://www.ukfinance.org.uk/policy-and-guidance/reports-and*https://www.ukfinance.org.uk/policy-and-guidance/reports-and-publications/annual-fraud-report-2022*publications/annual-fraud-report-2022*

FCA prior to serving financial advertisements on their sites.[3] However, the implementation has been slower than necessary, and the full impact has not materialised. While this demonstrates wider collaboration between industry, government, the tech and digital sector, further measures are needed to be undertaken to develop a comprehensive approach to stem the flow of fraud and scams.

- Despite controls that mitigate, filter and block paid for adverts by platforms, tens of thousands of malicious advertisements are served to consumers and commercial customers monthly. Platforms do not have the visibility of identifying adverts that pass through their filters, and victims do not capture the adverts' contents nor any images. Adverts are served to victims in various ways making it time consuming and ineffective to identify the specific version of an advert potentially served to a victim. Receiving intelligence of adverts that have reached victims is vital to the automating further mitigations for platforms. Criminal groups put a substantial amount of effort in to appearing legitimate to unsuspecting victims and businesses, often impersonating genuine organisations. Sectors that are impacted by rogue advertising have the potential to provide intelligence insights to the platforms to support the strengthening of their controls. Online platforms are receiving payments for the rogue adverts served, and so any funds received should be allocated to prevention and mitigation of fraud and scams.

- In addition to this, user generated contents for complex investment products such as non-fungible tokens (NFTs) are often not denoted with 'unverified' or 'not authorised' statuses. Following consumer testing, these verification statuses should be applied consistently by all platforms to raise awareness of the content's alleged fraudulent activity.

- There have been numerous occasions where regulated financial institutions are unable to get cloned/fraudulent contents removed, in the instance of impersonation of genuine brands or trusted entities such as the government (see quarantine hotels case study). The controls used prior to contents being posted need to be stronger and complemented by a sector-specific take down process and mandating direct access to moderators for regulated sectors e.g., financial services, investments, regulators for illegal content. Collaboration to support priority access could be supported via a number of initiatives such as the OFSG, Cyber Defence Alliance (CDA), CIFAS, Stop Scams UK or via an Ofcom issued SPOC list.

- It is also vital that feedback is shared by platforms on actions or non-actions taken to reduce the harms seen by consumers and to reinforce platform's policy guidelines.

Money Mules:

- A mule account is needed to move the funds acquired from each scam or fraud. Social media has long been the recruitment platform for criminal groups. In 2020 the Dedicated Card and Payment Crime Unit (DCPCU), working in collaboration with Facebook/Instagram, removed over 250 mule recruitment social media accounts. Each account will likely have thousands of followers and an even higher volume of views, which normalises this illicit content. Despite the proactive work by the DCPCU,

[3] https://www.techuk.org/resource/major-technology-companies-step-up-efforts-to-tackle-financial-fraud-andhttps://www.techuk.org/resource/major-technology-companies-step-up-efforts-to-tackle-financial-fraud-and-scam-adverts.htmlscam-adverts.html

a more accurate depiction of the scale of the money mule threat impacting the UK, and the harms associated can be ascertained from direct liaison with the NECC.

- Mule recruitment drives target social media users with sophisticated lifestyle accounts for marketing, using adverts and user generated posts, as well as trawling through related discussion groups, promising “easy ways to make money”, using images of cash and luxury items to entice young people. Once the mule recruiter has connected with their victims, they can then also be targeted via private messaging, with quick access or limited time offers. The user-generated posts and adverts on social media tend to showcase how a “normal” person managed to make large amounts of money through “simple” techniques or lucky investments, enabling them to retire and lead a luxury lifestyle, leading victims to falsely believe that they can achieve the same. Mules are also recruited via word of mouth and recommendations from existing recruits that have not yet discovered the reality of the activity they are partaking in. Where a mule wishes to extricate themselves from the criminal enterprise, they themselves can become targets of the criminal gang.

- The age range of those participating continues to drop, with CIFAS reporting that over 17,000 suspected money mule cases involved 21–30-year-olds in 2020, up five per cent on the previous year. The impact of being a money mule can have a long-term effect with on those recruited. Besides a criminal record, the individual could have their bank account closed and difficulty opening one elsewhere, and trouble obtaining mobile phone contracts or accessing credit in future.

- Given the prevalence of money mule recruitment online, the onus should be placed on platforms to proactively identify and take down money mule recruitment drives, as well as educating users on the dangers of becoming a mule. There should be promotion of reporting mechanisms, in particular for users that have encountered mule recruitment posts. Bilateral information sharing with sectors also provides a necessary mechanism to inform further take downs by platforms and inform the FS sector of targeted consumers. Social media companies should also have a responsibility to educate users of the dangers of becoming a mule, as well as having a responsibility to prevent such posts from appearing, take them down when identified and allow users to report mule recruitment posts.

Purchase scams:

- These scams usually involve the use of an online platform such as an auction website or social media. While many online platforms offer secure payment options, the criminal will persuade their victim to pay via a bank transfer instead, which provide no protection for the consumer. The risk then migrates from the platforms to the banks.

- Platforms’ organic contents are misused to advertise reduced price goods and services which are often purchased by unwitting consumers using faster payments. Criminals act as the middleman to purchase the goods or services with stolen cards details. As well as the victims of the fraudulent card purchase, this harms genuine businesses with the loss of merchandise and services as well as triggering the card scheme monitoring rules, putting firms at risk of losing payment facilities. The platforms are routinely misused to organically advertise the resale of the proceeds of crime [4]

- Separately to this, criminals also use online platforms to trade stolen card credentials or personal information, often using abbreviated terminology e.g. Fullz to mask their fraudulent activity. Proactive mitigation by online platforms for contents that have the most common terminologies used by criminals would allow trusted partners and law enforcement to refocus their resources. A two-way intelligence flow could robustly support platforms understanding of current and new terminology.

- Due to influencers having a large following, they are seen to be trustworthy and a reliable source of information. Legitimate influencer accounts are being taken over to direct followers to accounts associated with mule herding, or investment scams. The platforms are also accepting revenue for this, in particular for crypto. Proportionate controls that are proportionate to the risk of activity or contents are required.

- As mentioned previously, the FS industry is collating enabler data which will elucidate sources for the increasing volume of fraud and scams, this new insight will be shared with Ofcom once more mature. In the interim the analysis showing over two thirds of scams originate online has already been published by UK Finance. [5]

3. How do you currently assess the risk of harm to individuals in the UK from illegal content presented by your service?

12. With online and digital platforms such as search engines, social media sites, and instant messaging services serving as primary communication channels for a growing number of consumers and businesses, they will continue to be exploited for scams such as mule herding and hosting other criminality such as ghost brokering, investment scams and purchase scams. Systems and controls need to be introduced at the front end to manage this risk. This is especially pertinent for those platforms who have been identified as enabling larger volumes of illicit activity to be advertised on their sites. Certain scams and fraud types are more prevalent than others, as evidenced below.

Taken from the UK Finance annual Fraud Report 2022

Scam Type	Total volume of scam cases	Total value of victim losses (£m)
Purchase	99,733	64.1

[4] https://www.thisismoney.co.uk/money/beatthescammers/article-10368215/The-online-schools-scammers-avoidhttps://www.thisismoney.co.uk/money/beatthescammers/article-10368215/The-online-schools-scammers-avoid-victims.htmlvictims.html

[5] https://www.ukfinance.org.uk/press/press-releases/over-two-thirds-of-all-app-scams-start-online-new-ukhttps://www.ukfinance.org.uk/press/press-releases/over-two-thirds-of-all-app-scams-start-online-new-uk-finance-analysisfinance-analysis

Investment	12,074	171.7
Romance	3,270	30.9
Advance Fee	20,495	32.1
Invoice and Mandate	4,330	56.7
CEO	461	12.7
Impersonation: police/bank staff	29,406	137.3
Impersonation: other	26,227	77.5

13. The growth in online fraud does not just create significant financial losses, but also has a devastating emotional impact on victims. As outlined in the 2020 National Trading Standard's Well-Being Report, scam engagement has a negative impact on health and well-being, often undermining an individual's self-confidence and usefulness. [6] As onlineenabled fraud often involves victims being directly manipulated, or duped into making a fraudulent payment themselves, this can have a more damaging psychological impact than more traditional types of fraud.

14. Prevention needs to occur at the front end of the scam attack and needs to be proactively supported by the online platforms. Not all fraud and scams are reported, so there will also be a volume that remains unaccounted for. While banks repatriate victims of fraud, this is not necessarily universal for scam victims where assessment is conducted against the principles of the Contingent Reimbursement Model Code ('the Code'). Even if the customer is compensated in full by their financial provider, the organised criminal gangs that perpetrate these frauds still profit from the proceeds. Such monies can go on to fund illicit acts – such as terrorism, drug trafficking and people smuggling – that damage the fabric of our society, with fraud detailed as one of the indicative offences leading to money laundering.

15. Banks play a pivotal role in supporting and protecting people with a range of vulnerabilities. Supporting customers in vulnerable circumstances remains a priority for the FS sector and should also be extended to online platforms. Young people, in particular, are often less informed of the risks associated with engaging in online fraudulent activity such as money mule adverts. As a result, platforms should be using their online presence to provide tailored content to raise awareness and educate consumers of related fraud risks.

For purchase scams there is a need to:

[6] https://mycouncil.surreycc.gov.uk/documents/s77885/Item%204%20-%20Annex%20C%20-%20NTS%20Impact%20of%20call%20blockers%20on%20user%20well%20being%20Report%202020.pdf

- Strengthen online purchase controls or payment options and provide payment warnings to safeguard buyers from criminals.

- Keep an audit trail of messages and online adverts with the capability to formally confirm receipt or non-receipt of goods.

- Collaboration for protective payment mechanisms with payment industry, to ensure consumer protections in the online marketplace are akin to those of ecommerce.

- Devise a victim reporting mechanism for confirmed fraudulent activity complemented by an intelligence feedback loop to relevant parties i.e., banks, law enforcement.

- Use a trusted seller authentication process informed by ratings of activity and fraudulent reports to provide confidence to buyers.

- Publish fraud and scam statistics aligned to the FS sector's fraud reporting process for victims, as this could support tracking the impact of new controls.

- Mandate seller identification protocols using KYC controls i.e. ID checks, address and age verification to remove seller anonymity.

- Adopt consumer protection payment schemes that focus on the right behaviours – a template for this would be the controls enforced by the card's schemes on the ecommerce environment for card payments.

- Disseminate visible scam warnings across platforms to raise awareness and educate consumers of prevalent fraud risks in online marketplaces.

For investment scams there is a need for:

- Regular cadence of case study review and feedback to build a core understanding of the problem.

- Identification of suitable data items from across the ecosystem that can drive proactive prevention.

- Improved sharing of value-add data in as proactive a manner as is possible within the current legislation.

- Proactive reporting of known malicious advertisements including insight on users who have interacted with them prior to take down.

- Improved warnings relating to fraud and scams across platforms to build user knowledge and awareness.

4. What are your governance, accountability and decision-making structures for user and platform safety?

16. No response

5. What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements?

17. Consumers should be notified where they have either clicked on or seen “bad content” through a direct email from the platform or a splash page or videos outlining the reasons for removal of the post/account. The core messaging utilised should be consistent across all platforms.

18. The standards set within the terms of service needs to be communicated and made aware to consumers to reinforce the standard of content that should be seen. This can be communicated through digestible chunks, whether that’s through video content or recognition of good practices. There can be examples of good and rogue behaviours to bring the most common issues to the attention of new subscribers.

19. In the same way that algorithms continue to serve adverts following customer searches, AI should also be applied to serve customers and businesses with targeted educational content. For example, investment searches should lead to appropriate investment-related warnings being served. Where someone searches contents or clicks adverts, the platform algorithms naturally target that consumer with a high volume of similar contents. There needs to be a safety mechanism to mitigate rogue adverts bombarding consumers and in particular vulnerable consumers.

20. Additionally, a proactive and risk-based approach to warn consumers of fraudulent content, such as that relating to money mules, need to consistently be implemented and applied by all platforms. In the FS sector effective warnings are provided throughout a customer’s payment journey, to enable intervention prior to the conducting of transactions. Online platforms have previously demonstrated their ability to take urgent action to limit disinformation i.e. around Covid-19 vaccinations. This forms part of the Code of Practice on Disinformation which platform signatories including TikTok, Twitter, Meta, Microsoft and Google have committed to. [7] This aims to reduce manipulative behaviour used to spread disinformation (e.g., fake accounts and impersonation). The scope of this should be expanded to include harms relating to economic crime.

21. Platforms should further seek to build profiles and models of mules and ghost brokers using intelligence shared from sectors to intervene with education and awareness.

6. How do your terms of service or public policy statements treat illegal content? How are these terms of service maintained and how much resource is dedicated to this?

22. No response

7. What can providers of online services do to enhance the transparency, accessibility, ease of use and users’ awareness of their reporting and complaints mechanisms?

23. Reporting routes for victims should be consistent across all platforms with a central repository housing all reports/complaints. This should be implemented in a similar vein to that of the banking and finance industry which fully recognise the importance of tackling fraud and is already highly regulated with respect to this. The FCA is responsible for securing an appropriate degree of protection for consumers, and requiring firms to ensure communications are fair, clear, and not misleading.

24. The trends and intelligence that is gleaned from victim reporting and complaints should be expanded and utilised to further enhance the intelligence picture of the platform. This

[7] https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation

should be supplemented by other industries and regulated sectors where the harms materialise.

25. Attacks from different regions of the world and platforms should be shared in real time so as not to be repeated in our region. There needs to be a proactive approach for the safe sharing of attacks mechanisms within the industry and from impacted industries.

26. The lack of feedback loop with other impacted sectors means there remains a gap in understanding the impact of the harm, and where/why adverts bypass filters. Where platforms or other actors have machine learning tools and an ongoing feed of these rogue adverts this will help train and optimise this capability to be more effective. Online platforms sharing data on identified sellers of stolen or counterfeit goods would allow for the removal of access to banking from criminals who use it to process their illicit gains.

27. Furthermore, a mechanism that enables ease of weighted notifications, reports or complaints needs to be devised. A periodic aggregate report that demonstrates the volume of rogue content that is taken down between, automated rules, machine learning, human intervention, and external reports along with the typical SLAs for these segments would determine where there is a best vs more relaxed process. This would then inform the regulators’ guidance to the sector.

28. There is currently no awareness of the victim reporting processes, and there is a missed opportunity to understand victims' profile to enhance messaging or mitigation to targeted segments or the demographics of victims. Insights of the victims identified as being targeted on the online platforms could also help other impacted sectors take forward mitigating activity. Where the platforms could group the reporting for different offences, there could be opportunities to create machine learning models for the different types of economic crimes that are occurring.

29. The highest standards to protect online consumers should be enforced and monitored by Ofcom. As mentioned previously, the origin of fraud and scams via social media platforms demonstrates the ineffectiveness of platform’s self-regulation. All new products or platforms introduced by current providers should be subject to measures imposed by Ofcom. Furthermore, the use of advertising to engage potential victims is a live issue. Within the financial services sector we regularly see criminality migrate further along the chain to the next weakest link; there needs to be a strong regulatory expectation and mandates to mitigate this live issue; otherwise, there will be repeat attacks and ongoing iterative issues across the ecosystem.

30. The due diligence approach used within the financial sector (see current-guidance) illustrates how a regulated gatekeeper sector is required to manage risks to the legitimate economy. Other sectors that are being exploited to introduce risk of economic crime into the system should be required to take their own equivalent measures.

31. There should be obligatory reporting of fraud or compromises to allow analysis of typologies to determine common gaps. For example, if it is found to be commonplace that lower-tier services have a greater volume of misuse due to disparity in the deployment of due diligence controls, the regulator could make these controls a condition of licence.

32. The reporting mechanisms should be consistent across all platforms and could be highlighted using a single universal icon e.g., the microphone symbol, which is universally

known by all regardless of platform, operating system, or device type. This would help streamline the consumer experience across platforms.

33. Users, or trusted individuals acting on their behalf, should be able to report scams/mules/recruiters, filter adverts and block sites/pages. Reporting routes for victims, or trusted individuals acting on their behalf, should be consistent across all platforms with a central repository housing all reports/complaints. As well as thresholds and algorithms based on the reports that automate escalation for review and/or takedown or harmful contents.

8. If your service has reporting or flagging mechanisms in place for illegal content, or users who post illegal content, how are these processes designed and maintained?

34. During the payment cycle the financial sector is often unaware of the initial scam advert that a victim encounters, and currently does not receive any intelligence that the victim has been targeted until long after the transaction on their account has taken place.

35. Currently there appears to be disproportionate weighting on other sectors to utilise additional services e.g., trusted partners and law enforcement, to police platform contents and engage for subsequent take downs. The DCPCU, funded by and working on behalf of the banking industry, ascertained trusted partner status with Meta, Instagram, TikTok and Snapchat. Working closely with platforms, they have successfully taken down 1,032 accounts between 2020-22 – these accounts had thousands of followers and views which has a normalising effect on their content. Platforms should instead proactively police their own sites and source additional intelligence and trend insights from impacted sectors via sector specific flagging mechanisms. All users that view and follow such contents should be given educational material to ensure they are aware such content is unacceptable.

36. There are now tools to scan advertising images and leverage machine learning capability; these tools require training to distinguish both genuine and harmful content examples. Sharing of example images across different platform types and global regions would optimise the mitigation of repeat attacks against differing parts of the digital advertising ecosystem. Analysis of the adverts source would also identify weaknesses of controls within the advertising ecosystem.

37. Despite the process/mechanisms that platforms have in place, there is still a volume of bad traffic that bypass their filters. A regular and constant intelligence feed of automated and user generated adverts that reach consumers is required, complemented by other sectors, for platforms to be more informed on the mitigation of bad content and the use of terminology used by criminals. This can be facilitated through existing intermediary bodies which have a central intelligence function and act as an agent of the FS sector i.e. Cifas, CDA, OFSG via UK Finance.

<u>Example of collaboration</u>:

- By sharing best practices across the FS sector, an online platform has demonstrated mitigation techniques can be enhanced. While there are filtering techniques used to mitigate rogue adverts by online platforms, new intelligence capabilities need to be developed to determine if the URL is pointing to fraudulent content. These are often used in fake adverts, and relates to the below:

    - Phishing ○ Cryptojacking
    - Malware Binary Infrastructure URLs
    - Web shells

- Fake shops
- Scams

38. The intelligence feeds of some tools identify the underlying code of URLs, as platforms can face substantive cloaking by criminals. URLs on their own are often a bit of a blunt instrument and further analysis is required.

9. If your service has a complaints mechanism in place, how are these processes designed and maintained?

39. An effective complaint classification system consistent across all platforms needs to be devised in order to assess the efficacy of new controls. The alignment of classification with definitions used by the FS sector for fraud and scams would be beneficial to determine impact of control changes. The FCA provides detailed rules for the handling of complaints, which also requires the regulated sector to analyse the root cause of all complaints and remedy any systemic or recurring issues.

40. Furthermore, where a repeat offender has been identified this needs to be shared across all platforms, to inform enhanced monitoring and awareness to prevent repeat attacks on consumers across multiple platforms.

10. What action does your service take in response to reports or complaints?

41. Where a scam or fraudulent activity has been reported, the FS sector aims to escalate the account/post to the relevant platform for take down. Additionally, proactive threat hunting is undertaken by the FS sector to uncover OCGs and work with law enforcement.

Below are a couple of examples of experiences from our members:

Bank A

Over the last year, we have reported eight brand infringement social media sites to Meta via our take down provider. Only five (62.5%) were removed within 24 hours with the remaining taking nearly 72 hours to be removed. We are aware of several unauthorised xxxxxxxxx bank pages and have attempted to have them removed. Communication has been challenging without a point of contact and despite providing evidence we have not been able to have these pages removed. This can cause confusion for customers and has a detrimental impact on our brand reputation.

The potential harm for contents that remain up for a period of time needs to be captured, and the volume of views or follows that have occurred from the time of report to the time of removal needs to be tracked.

Bank B

We see a noticeable difference between Meta and other platforms when dealing with requests to remove fraudulent content.

The issues have arisen mainly with Instagram where there were profiles impersonating named executives of Bank of xxxxxxxx. It is worth making the distinction between those types of issues and other Meta profiles that impersonated ‘bank b’ as a corporate identity – in those ‘corporate’ cases, the profiles were removed reasonably quickly once they were notified.

The issues with getting Instagram executive impersonation profiles removed is as follows:

- No mechanism to report fraudulent content other than through the standard in-app reporting portal (we have direct security contacts for other platforms).

- Poor responses from automated addresses, with no ability to reply or follow up, e.g. from last month *“We have fewer people available to review your request due to the coronavirus (COVID-19) outbreak. We’re only able to review requests for the most urgent cases. This means we can’t review your case right now. Please try again at a later date.*

- Automated responses to notifications of obviously fraudulent profiles saying things such as the profile did not breach any community standards on the platform.

- No acknowledgement of or allowance for the fact that fraudulent content could be reported by a bank’s security team or an authorised third party, e.g. some recent responses to reports of fraudulent profiles impersonating senior executives have included, “*We can't help you with your request until we receive an ID or other document that we can use to confirm that you're the owner of this account. The document should include your name, photo and date of birth that matches the info on this profile.”;* and “*Thanks for your response. To confirm that you're the owner of this account, you'll need to reply to this message and attach a photo of yourself holding your government-issued photo ID.”*

- Little or no engagement on wider basis on security / fraud issues or areas of potential cooperation.

42. There is a need for platforms to leverage consumer reports to aid profiling for insights on the victim or targeted demographics, terminology of user generated contents, of victims’ and the evolving trends of criminal approaches.

43. The FCA’s focus is on our financial promotion's regime which centres around whether something constitutes an inducement to consumers to invest in a regulated product in breach of s21 FSMA. This can include paid ads as well as organic content. Agreeing a publication of stats between the FCA and Ofcom relating to rogue promotions would be beneficial to demonstrating the impacts of new controls.

11. Could improvements be made to content moderation to deliver greater protection for users, without unduly restricting user activity? If so, what?

44. Content moderation by platforms is essential to the safeguarding of online consumers. As mentioned previously, two-way information and data sharing will help inform the threat picture and accelerate proactive mitigation.

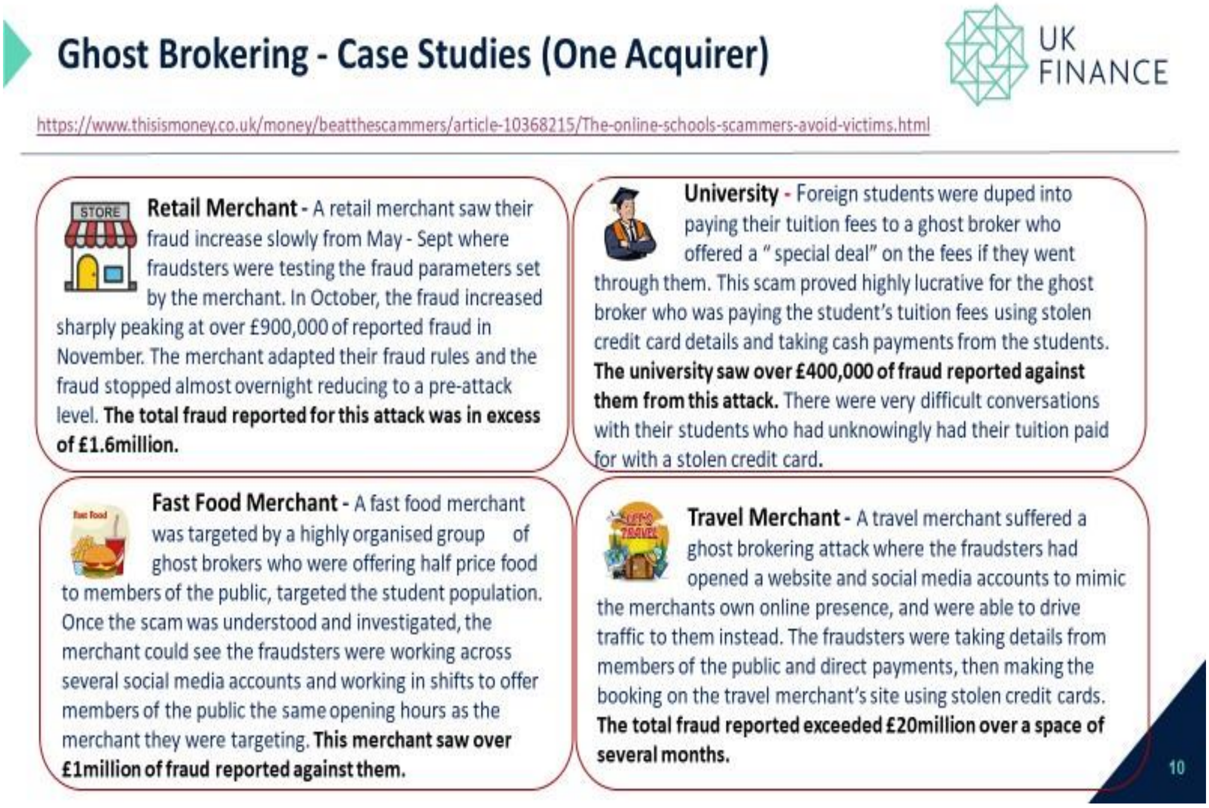
.

45 Platforms should be required to conduct risk assessments to identify the risks most prevalent to them for different activities, and highlight the mitigations they have in place, with particular focus on the conducting of financial transactions. Platforms are often used as the gateway to conduct initial conversations before messaging is moved off from the platform as the interface.

46. Posts/adverts pertaining to ghost brokering/fake investments and impersonation scams should be screened for by social media platforms to accelerate the take down of contents and disruption of criminals. All accounts advertising fraudulent activity should be proactively blocked and deterred from creating further accounts on any platform. **Ghost brokering case study**

- This is where the criminals advertise goods or services at a significantly reduced price, the consumers pay via a direct bank transfer to the criminal and the criminal uses payment details such as stolen card data to acquire the goods from the merchants.

- The victims see unauthorised card transactions on their statements and the merchants lose both the goods and the value of the transaction. In addition to this, prolific attacks can put the merchant into monitoring programs of the cards schemes and potentially at risk of losing legitimate payment facilities.



Ghost Brokering - Case Studies (One Acquirer)

https://www.thisismoney.co.uk/money/beatthescammers/article-10368215/The-online-schools-scammers-avoid-victims.html

Retail Merchant - A retail merchant saw their fraud increase slowly from May - Sept where fraudsters were testing the fraud parameters set by the merchant. In October, the fraud increased sharply peaking at over £900,000 of reported fraud in November. The merchant adapted their fraud rules and the fraud stopped almost overnight reducing to a pre-attack level. **The total fraud reported for this attack was in excess of £1.6million.**

University - Foreign students were duped into paying their tuition fees to a ghost broker who offered a " special deal" on the fees if they went through them. This scam proved highly lucrative for the ghost broker who was paying the student's tuition fees using stolen credit card details and taking cash payments from the students. **The university saw over £400,000 of fraud reported against them from this attack.** There were very difficult conversations with their students who had unknowingly had their tuition paid for with a stolen credit card.

Fast Food Merchant - A fast food merchant was targeted by a highly organised group of ghost brokers who were offering half price food to members of the public, targeted the student population. Once the scam was understood and investigated, the merchant could see the fraudsters were working across several social media accounts and working in shifts to offer members of the public the same opening hours as the merchant they were targeting. **This merchant saw over £1million of fraud reported against them.**

Travel Merchant - A travel merchant suffered a ghost brokering attack where the fraudsters had opened a website and social media accounts to mimic the merchants own online presence, and were able to drive traffic to them instead. The fraudsters were taking details from members of the public and direct payments, then making the booking on the travel merchant's site using stolen credit cards. **The total fraud reported exceeded £20million over a space of several months.**

47. Having worked with one platform there is an increased understanding of the tools used by the FS sector, and the scope to leverage existing capabilities and measures to protect against a multitude of harms.

48. Account onboarding underpinned by due diligence from either human moderators or automation should be mandatory for all users setting up new accounts. There is specific focus on the lack of onboarding due diligence for online marketplace, advertisers or user

generated contents relating to purchases, as these drive the highest volume and over half of all APP scams.

12. What automated moderation systems do you have in place around illegal content?

49. Currently we are aware natural language monitoring is utilised on some platforms, however the various covert terminology used by criminals needs to be considered for inclusion as an intelligence source. This will need to be updated to denote changes as criminal enterprises evolve their techniques.

50. The proposed Digital Identity and Attributes Trust Framework will see further introduction of additional digital identity service providers in the UK market, besides the existing Yoti and Post Office EasyID and HooYu Limited, among others. Online platforms and marketplaces should be required to verify the identity of new accounts – this could be facilitated through the use of trusted identity service providers under the trust framework.

51. Despite heavy focus on the impact of fraudulent activity on buyers, protection also needs to lend itself to sellers with mechanisms devised to report scams as a seller. Some sellers are targeted with malicious links to infect their devices.

13. How do you use human moderators to identify and assess illegal content?

52. The use of human moderators is essential to identifying new trends and assessing illegal content, and this enables informing tools as criminal techniques evolve beyond those captured using automation.

53. As mentioned, the DCPCU is a trusted partner of several online platforms. They review the platforms for core terminology being used by criminals to locate accounts that are focused on criminal activities. This is then notified to the platforms for urgent take down. While this is positive, the terminology used can remain consistent for substantive periods, as such the platforms could proactively remove the contents without the need for additional reports from law enforcement.

54. There are currently two reporting routes; one for trusted partners and the other for the regulated sector, regulators and consumers. Each should have a weighted priority when reporting, to support accelerated take downs for higher harm contents. There should also be an escalation process for opaque contents that are repeatedly flagged by multiple sources.

55. When looking at the higher end of criminality, where lifestyle accounts often use images such as flash cars or large sums of cash to recruit money mules, the need for human moderators is far greater to assess context and implications. Often the platforms are used as the gateway for further conversations through direct messaging on the platform or via alternative encrypted channels.

56. There are watch lists produced by the FCA to help prevent consumers that wish to invest from becoming victims, however some of the named individuals on this list have made it on to the platforms with verified status. This would suggest human intervention would be more accurate than an automated processes. Often lifestyle accounts are use which use high-value goods or cash to entice the public rather than substantive written contents.

57. Furthermore, enhancements are required for the FCA list to be in a more digestible format and provide the necessary unique indicators and information for it to be more widely used by all sectors.

14 How are sanctions or restrictions around access (including to both the service and to particular content) applied by providers of online services?

.

58. No response

15. In what instances is illegal content removed from your service

59. The threat landscape for other sectors needs to be shared to understand the financial crime issues that are of prevalence, and their associated harms. As mentioned earlier, twoway information and intelligence sharing is needed, particularly relating to language used by criminals. This is integral to underpinning the identification of fraud and scams and to its subsequent mitigation. The financial services industry is subject to legislation, regulations and codes of conduct which create the need for banks to take action against fraud and scams. We believe it is essential that the online sector should be subject to similar regulation and legislation in order to protect consumers and businesses,

60. Where images and slang terminology are used to generate conversations relating to fraudulent activity, such as the recruitment of money mules or ghost brokering, challenges may arise in terms of the removal of such content. This also extends to the use of temporary content i.e., Instagram stories/Snapchat which can be used as a mechanism to promote/advertise fraudulent activity across a short period of time before disappearing. In these cases, regular due diligence across the platform needs to be undertaken to prevent consumer harms. There is also a need for aftercare to engage those that have seen the contents and been targeted as victims.

61. As mentioned previously, criminals have repeatedly hijacked influencer accounts to promote their fraudulent activity, which can be evidenced in the 2020 bitcoin scam where the accounts of Joe Biden and Bill Gates were hacked to solicit bitcoin transactions. Alternatively, genuine influencers may also promote scams such as investment unknowingly, hence calling for enhanced due diligence to be conducted on all accounts advertising products and services. The hijacking of accounts also lends itself to genuine account takeover, where the name of the existing account is then amended.

16. Do you use other tools to reduce the visibility and impact of illegal content?

62. Currently natural language is used to monitor for nefarious activity such as terrorism, if trained to detect for scams and fraud terminology this could improve platforms' monitoring capability.

63. Specifically, to acquirers (merchant payment service providers) in the FS sector a web crawling system is currently used as part of ongoing due diligence to monitor content across various websites. Given the need for enhanced monitoring across social media platforms, search engines and online marketplaces, learnings can be shared to identify any potential opportunities for implementation of a similar tool across the tech and digital sector. This would help alleviate other sectors and law enforcement to regularly monitor social media platforms and online marketplaces to identify fraudulent content.

64. Following Google's announcement to implement a verification policy for FCA-approved advertisers who wish to run ads for financial products and services, statistics from one member have shown a decrease from 6.75 per cent to 5.53 per cent in google-enabled claims. This is a positive change and one that should be implemented across all search engines and online platforms as quickly as possible.

17. What other sanctions or disincentives do you employ against users who post illegal content?

65. An updated list of users who have been sanctioned/disincentivised should be shared regularly across all platforms to mitigate the potential of user migration. This should include images and terminology or context to aid mitigation where bad actors are adjusting their tactics.

66. Furthermore, a migration of harms from SMS to WhatsApp and Apple iMessage has been identified following increased controls by telecoms. One member reported the average WhatsApp scam claim to be £1.2K between November 2021 to April 2022. This mirrors the previous levels of attack experienced by other regions of the world and reinforces the need for all sectors to implement equivalent controls, share best practices and information across regions for effective mitigation. The real-time nature of detection and advanced techniques with external learning feeds will enable a more effective mitigation of consumer harms and proactively disrupt repeat attacks.

67. Organised criminality will migrate to exploit vulnerabilities found in other regions and further upstream to circumvent the controls deployed within the UK.

18. Are there any functionalities or design features which evidence suggests can effectively prevent harm, and could or should be deployed more widely by industry?

68. Given that a number of the economic crime harms will surface in the financial services sector, having bypassed platforms' filtering controls, there is a need to for two-way information and data sharing across sectors and regions to inform the threat picture and accelerate mitigation. Our members have provided a summary of real case studies which helps demonstrate the range of attacks seen by consumers in Appendix 1 Case Studies.

69. The enhancement of granular economic crime categorisations could support feedback loops between sectors, to enhance the existing detection and take down of adverts linked to scams. Secondly, the FS sector can protect potential victims where data has been harvested. This approach would aid transparency of the harms and impact of mitigation activity.

70. For the crime types, there is a need to create a mechanism for automated intelligence sharing to support automated models that detect illicit activity. Examples of this would be to create separate models to detect mule herding/coordinating and also basic mule accounts. The treatment of these two segments would differ in that the herders could be pursued via law enforcement and the mules could be served additional education materials or warnings. A consistent approach for information sharing such as standardised dissemination or fraud trends and methodologies could drive improvements for automated fraud prevention tools as trending key words are added to fuzzy logic.

71. The sectors that are being impersonated/cloned or receiving complaints in relation to scams have sight of the adverts that surpass the filtering and initiatives, but the lack of feedback loop and two-way data sharing is a significant gap to mitigating fraudulent adverts and protecting victims from subsequent harms having clicked through an advert.

72. As a minimum, the sector is not sharing intelligence and examples at volume or pace for contents that has evaded detection systems to protect the public from criminals that repeat attacks via different platforms. This should be a priority for online marketplaces, especially where a high fraud rate for purchase scams has been identified.

73 The utilisation of a baseline measure across the differing platforms and scam types will help to highlight disparities between platforms. Learnings can subsequently be taken from those platforms where issues have not been suffered.

.

74. Additionally, the fraud rate for transactions should be calculated to help understand context. This could inform the risk assessment based on activity on the platform. The consumer reports could help to access the impact of differ contents and scam or fraud typologies. The sectors impacted could also contribute to a sample evaluation to help inform and support a projection of impacts.

20. How do you support the safety and wellbeing of your users as regards illegal content?

Once a consumer has been exposed to illegal content, a proactive approach is required by platforms and marketplaces to offer consistent advice on keeping safe from fraud and scams and signposting to after-care as follow-up. This could include serving video content that cannot be exited until watched from start to finish, with live checks to support reasonable assurance that the recipient has watched the full contents.

An example of good practice by Ebay is seen below:

> *It looks like you may have been in touch with sm-749372 and we want to let you know that some security concerns have come up with this account. We recommend taking some precautions to make sure your account stays secure.*
>
> *Here's what you should do*
>
> - *Stop all communication with sm-749372.*
>   - ✦ *Don't respond to their offers to buy or sell an item.*
>   - ✦ *Don't respond to requests to change your order or delivery information.* ✦ *Don't click on or copy and paste any links they send you.*
> - *If you sent an item to sm-749372, contact the carrier to have it returned to you.*
> - *If you sent payment outside of eBay, contact your payment service provider to alert them of a potentially fraudulent charge.*
>
> ***You may also want to***
>
> - *Update your password*
> - *Change your secret questions*
> - *Forward any suspicious eBay emails to spoof@ebay.com*
>
> *We're here to help. If you have any questions or concerns, just get in touch.*
>
> ***Help & Contact***
>
> *We appreciate you being part of the eBay community.*
>
> *Thanks,*
>
> *eBay*
>
> *Please don't reply to this message. It was sent from an address that doesn't accept incoming email.*

76. While the approach by Ebay in relation to customer education is seen as proactive, the mechanism in which it's delivered (e.g., Email vs In-app) needs to be considered to ensure the right approach is taken. Online platforms should seek to leverage their own In-app functions to educate customers on fraudulent activity and the risks associated with engaging, as spoofed impersonation emails are a common approach for phishing attacks. Directing consumers to webpages rather than leveraging In-App contents will result in a drop off of consumers ingesting education materials.

21. How do you mitigate any risks posed by the design of algorithms that support the function of your service (e.g. search engines, or social and content recommender systems), with reference to illegal content specifically?

77. The FS sector looks after the victims of scams, gaining insights on the criminal methodology and actively threat hunts across the online platforms to provide strong customer awareness campaigns. The sector also funds the DCPCU to pursue organised criminal gangs targeting the financial sector.

78. There are a number of common issues that require a proactive approach from the platforms to mitigate:

- **Compound harms example**: A brand impersonation of an investment company will often also result in a retail banking customer becoming the victim of an investment scam as funds will often come from their personal accounts. In this scenario there are two separate businesses handling the impact of a single rogue advert; and a customer that becomes a victim.

- **Money laundering example**: Due to influencers having a large following, they are seen to be trustworthy and a reliable source of information. Legitimate influencer accounts are being taken over to direct followers to accounts associated with mule herding, or investment scams.

- It was noted by a member researching rogue adverts, where a customer clicks through to a scam advert the algorithms (that are not yet adjusted to mitigate the latest rogue adverts) will continuously serve similar scam adverts, which compounds the potential that they will eventually become a victim of a scam.

Search engines

- Controls or Ts&Cs to manage the contents served need to be enhanced to ensure the public are confident when clicking on contents from a regulated sector vs unverified. The mitigation of rogue contents such as redirecting to a splash page that denotes scam contents were removed for a set period of time such as 24-72hrs.

Online marketplaces

- Purchase scams disproportionate in terms of volume of scams identified. For purchase scams there is a need to:

  ○ Strengthen online purchases processes and provide payment warnings to safeguard buyers from criminals. ○ Keep an audit trail of messages and online adverts with the capability to confirm receipt or non-receipt of goods.
  ○ Collaborate on protective payment mechanisms with payment industry, to ensure consumer protections in the online marketplace are akin to those of ecommerce. ○ Report fraudulent activity centrally. ○ Develop trusted seller verification. ○ Publish fraud statistics.

○ Ensure the identification of sellers. ○ Adopt consumer protection payment schemes to ensure fair outcomes.

○ Develop scam warnings tailored to the consumer demographics or potential scam type. **Online platforms**

- The FS sector has partaken in various initiatives to mitigate the risks posed by fraud and scams, including investment fraud. As an industry, regular intelligence calls are held to aid mitigation, and we participate in an authorised list which provides the genuine account details of participating investment firms that reduces blocks on transactions.

- In addition to this, we have started exploring the intelligence our sector holds in relation to rogue adverts to lessen the harms to victims.

22. What age assurance and age verification technologies are available to platforms, and what is the impact and cost of using them?

79. As previously mentioned, the proposed Digital Identity and Attributes Trust Framework will see further introduction of digital identity service providers in the UK market – many of whom are likely to provide age verification services.

80. The 'Age Estimation' system from Yoti was originally developed to age adults. It works by comparing the user's facial features to images of millions of people, for individuals aged 6–18, the software has a margin of error of just 1.5 years.

81. Instagram asks for age before showing you posts that are marked as sensitive. It's been blurring sensitive content for years, but now if you want to see contents the user's birthday will need to be on file with the platform. [8]

82. Where platforms can segment users based on age, additional controls can be applied to restrict direct communications teenagers and unknown adult users. [9]

23. Can you identify factors which might indicate that a service is likely to attract child users?

83. No response

24. Does your service use any age assurance or age verification tools or related technologies to verify or estimate the age of users?

84. Analysis from the FS sector has identified a growth in underage child money mules, an age bracket that requires additional safeguarding due to their ease of being influenced and heightened vulnerability. The FS sector is currently exploring additional reporting frameworks for parents/carers/teachers to report instances of suspicions or acts of money muling.

85. Informed by a user's search history, online platforms should seek to build profiles specifically for those who are deemed to be more vulnerable. The detection of those more susceptible to money muling can allow for more targeted consumer messaging and a proactive approach to undertake further mitigation.

[8] https://siliconangle.com/2017/03/24/instagram-explains-started-censoring-sensitive-content/

[9] https://www.theverge.com/2021/3/16/22333580/instagram-bans-adults-messaging-teens-safety-notice-prompt

86. The use of the digital identity framework will be integral to verifying the age of users as this is a gap across all platforms and should be mandated within a platform's sign-up process. The current lack of age verification enables consumers to gain access to platforms by changing their birth year. Search history, natural language and facial scan monitoring should also be used as an indicator to verify age ranges.

25. If it is not possible for children to access your service, or a part of it, how do you ensure this?

87. No response

26. What information do you have about the age of your users?

88. The FS sector can obtain victim data, particularly in cases where a child uses their parent's card details to purchase goods/services online. The current trends being observed are as follows:

- Our members are increasingly seeing younger victims targeted by fraud and scams.
- Mule recruitment posts and online marketplaces are specifically targeting a younger demographic. Tik Tok is a good example of a platform that is being used for recruitment and naturally has a younger audience.
- These individuals are arguably vulnerable and need and deserve better protection.

89. Verification of users informed by patterns in search history is necessary for content moderation, the management of which needs careful consideration before it can be implemented. Similarities can be drawn from cinema ratings where different ratings are used to communicate the level of content shown and help to inform if movies are appropriate for various age groups.

90. Self-certification or declaration of age is easy to override where a child claims to be over 16 or 18. A notification across platforms for vulnerable or young users could strengthen prevention controls across the online platforms.

91. The gambling sector has controls for underage gambling. Leveraging the learnings from this sector could help develop stronger practices to protect young individuals more robustly.

27. For purposes of transparency, what type of information is useful/not useful? Why?

92. Given the scale and volume of users on platforms and marketplaces, an aggregated transparency report by platforms should be produced on a quarterly basis. This should provide a holistic view of the measures and controls implemented by individual platforms and showcase not only the figures pertaining to take downs but also the scam adverts that were not captured. The report should not only focus on the proactive take down by platforms but also those that are respective and their relevant timelines for take down.

93. It would be beneficial for the report to include information pertaining to the number of scam adverts/posts taken down and the interaction with the fraudulent post before take down, and how long the advert/content was live. Further, this report should be aligned with industry categorisations of scam types, and their associated definitions, to ensure consistency.

94. Through its own transparency report, Ofcom should make available the following information: volume of complaints, volume of take downs (broken down by scam type), volume of people who saw the content, average time for removal. This could be produced in

a similar manner to the annual Fraud Report produced by UK Finance. Through its publication, platforms or marketplaces not in line with best practices could be identified and relevant guidance be produced to raise standards.

95. We also believe that there will be a subset of data that the platforms and marketplaces will not know i.e., how many scams bypass their filters. The FS sector will be able to provide data regarding the enablers of the scams that do still ultimately take place and result in the payment from a victim to a scammer. This data could provide deep insight as to the effectiveness of the measures that are being put in place by the platforms to detect and prevent fraudulent content, highlighting and evidencing areas where more needs to be done. We would therefore encourage Ofcom to work with the PSR, which is directing PSPs to publish data regarding their scams figures and reimbursement rates, to also require PSPs to publish their scams enabler data. This would require the FS industry to publish the data in a consistent way, resulting in it being more effective and reliable.

28. Other than those in this document, are you aware of other measures available for mitigating risk and harm from illegal content?

96. The proactive drive of collaborative culture between the segments of the ecosystem will accelerate the effectiveness of achieving the aims set out. The sharing of real-time intelligence and new typologies are critical to prevent attacks being replicated across the sector. Below are some approaches used in the financial sector to limit repeat attacks and reduce the number of potential victims', which could be considered:

- **Formal typology publications**: Publication of new typologies on a periodic basis; to ensure the guidance and expectations do not become dated as the criminals evolve their tactics.

- **Within industry**: In the financial services sector there is a culture of sharing intelligence and information to prevent repeat attacks by criminals across several banks. Example: UK Finance holds routine weekly intelligence sharing calls on behalf of our industry, which often summarise the latest techniques the criminals are deploying in their attacks on consumers or infrastructure. There is an industry strategic threat management process to surface new and emergent attacks that are scalable. Our sector also shares typologies during bilateral meetings with other sectors and regulators to accelerate the dissemination of criminal approaches to circumventing controls.

- **Cross sectoral**: Building a comprehensive regulatory approach to tackling online fraud requires cross-sectoral collaboration. It should be encouraged by Ofcom for those regulated in scope of 'the Bill' to participate in cross-sector initiatives driven by organisations such as CIFAS, UK Finance, OFSG and Stop Scams UK and maintain regular engagement to identify fraud threats and areas for mitigation.

97. In the banking sector, funds that are moved out of one bank will often become another bank's money mule used to launder the proceeds. This has encouraged strong collaboration across our sector to prevent further losses and consumer harms. For collaboration within and across sectors, there is often a need to achieve comfort and trust when establishing new intelligence sharing activities. However, there can be a reticence to engage with peers that are competitors. As such, often there can be differences within business risk appetites to share information and data. The oversight authority could mitigate this often-underestimated challenge, working in partnership with the ICO or Ofcom may help address concerns in advance, to ensure data sharing can be mandated. Leveraging existing industry led initiatives such as the Trustworthy Accountability Group (TAG), whose focus is

to fight criminal activity throughout the digital advertising supply chain could also build capability.

98. Platforms need to not only be taking responsibility to mitigate the issues but also accept liability to contribute to solving the problem and refunding where mitigations have failed to stop scams. This could be respective to the proportionality or scale of scam enablement seen across their platform and should be reviewed quarterly or annually. The pool of funds dedicated to supporting the technology that can mitigate/ prevent fraud and scams should also be reassessed quarterly or annually.

99. Platforms should also place the greatest importance on the communication of education and awareness with their customers, particularly relating to aftercare for impacted users that have encountered illegal or harmful content. This would raise awareness of the issues to consumers, demonstrate that the platforms are trying to protect their customers and raise the standards consumers expect to encounter online.

100. We recommend broadening the burden of financial liability to include the in-scope actors and a need for a strengthened range of sanctions powers. This could include a policed fine system whereby the firm that continues to perpetuate harmful online advertising is given an increasing fine over and above sharing liability on individual cases. This would create a backstop to help prevent cases that never make it to an actual fraudulent payment, helping to tackle fraud at source.