Meta – Written evidence (FDF0052)

Meta welcomes the committee's inquiry into the Fraud Act 2006 and Digital Fraud and would like to thank the committee for inviting us to contribute evidence on such an important topic. We have focussed our submission on our own efforts to tackle fraud on our platforms.

Meta's overriding priority is the safety of our users and we therefore take a zero tolerance approach to fraud on our platforms. By its very nature fraud is adversarial and hard to spot and the perpetrators of fraud are continually searching for ways to subvert the rules, processes and safeguards we put in place to protect our users. Just as it is unlikely that fraud will ever be eradicated in society at large, it is unlikely we will ever be able to completely eradicate it online. Nonetheless Meta is committed to doing all we can to prevent fraudulent activity on our platforms wherever we can.

There are currently around 40,000 people working on safety and security at Meta, around half of those directly review content and to date we've invested more than $13bn in teams and technology in this area. We have a team of highly trained experts solely focused on identifying fraud and building tools to counter this kind of activity, which are used to help catch suspicious activity at various points of interaction on the site, block accounts used for fraudulent purposes, and remove bad actors.

This response details the measures we take as an organisation, and in partnership with other organisations to combat fraudulent activity.

Community standards

It is directly in our interests to do all we can to combat fraud on our platforms. Failure to do so will expose our users to risk, severely degrade the experience of using our platforms for users and make them an unattractive place for brands and businesses to advertise.

Meta has a set of strict **Advertising Policies, Community Standards** and **Community Guidelines**, which govern what is and is not allowed in advertising and non-paid (organic) surfaces on Facebook and Instagram. These policies are designed to help ensure a positive experience for the people and businesses that use our platform, and are developed in consultation with experts. These rules set an important benchmark which enable us to set a high standard and to establish a basis for removing accounts or content which fall below that standard. In particular, our Advertising policies prohibit ads that contain low quality content or misleading claims. Our Unacceptable Business Practices ads policy does not allow the promotion of *"...products, services, schemes or offers using deceptive or misleading practices, including those meant to scam people out of money or personal information"*.

These policies are able to adapt and change if needed and our fraud and deception policy is updated regularly to better capture prohibited behaviours/content and deal with new emerging trends.

All content, including ads, can be reported to us by people, businesses or other organisations. We also specifically support a business's ability to identify and report instances where their assets are being used in inappropriate ways (such as business impersonation).

Where we believe anyone has violated our terms, standards, and policies we take action and use a range of tools to **enforce our policies,** either via proactive automated systems and/or reactive methods.

Fraud perpetrated through organic content

We deploy a combination of proactive detection and reactive action to disrupt bad actors on our platforms. This includes using our Artificial Intelligence (AI) systems to proactively detect suspicious activity. We focus our attention on behaviours rather than on content, as while the content of these scams changes frequently, the modus operandi of the bad actor typically remains the same.

To do this we analyse and incorporate typical behaviours you would associate with scams, such as bulk friend requests to certain demographics, poor feedback from users and likely fake profile characteristics. When we identify violating content, we will remove it. Where we identify profiles that persistently post violating content we will remove them from the platform while steps are taken to confirm their authenticity. If they are unable to prove their authenticity they are removed.

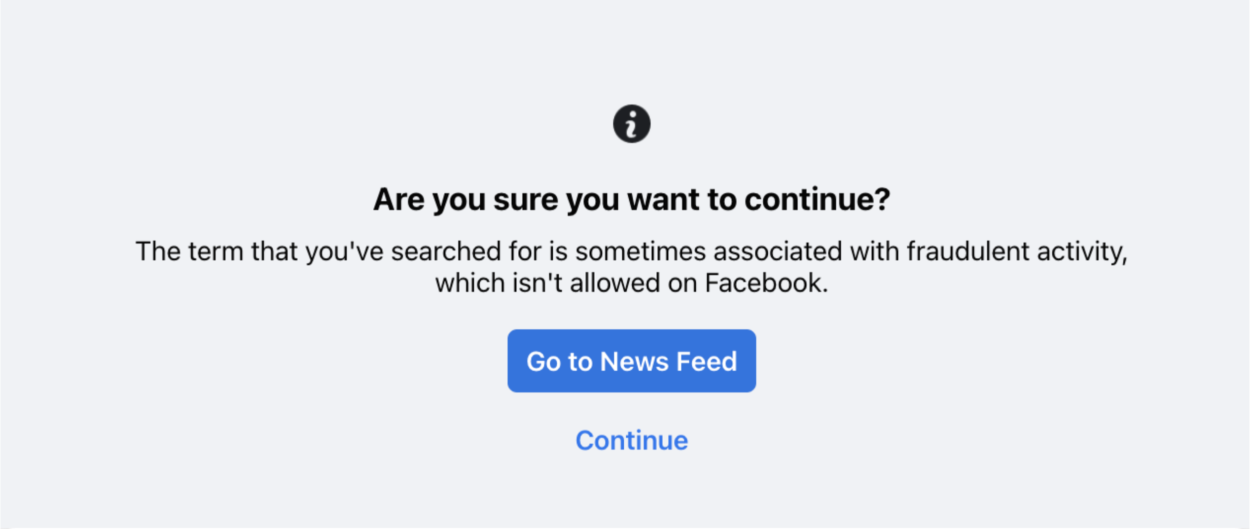
Where our systems are near-certain that content or profiles are violating (because they possess the signals we associate with a scam to a high degree of confidence) they will immediately be automatically removed. Where less certain, content may be prioritised for our moderation teams to review.

Our aim is to catch bad actors proactively, as early as possible, before they have a chance to engage with users. When someone looks to create a page or profile we will use our AI to check for signs they are being created by real people and not automated bots. Fraud and scams are often carried out through fake accounts. To combat this we have developed technology which enables us to find and remove fake accounts. In Q4 2021 we removed 1.7bn fake accounts from Facebook, 99.9% before they were reported to us.

It's important to note that when such accounts are removed, any associated content, posts, active message threads, etc or where these entities are a single owner of a Page or Group, these will also be removed.

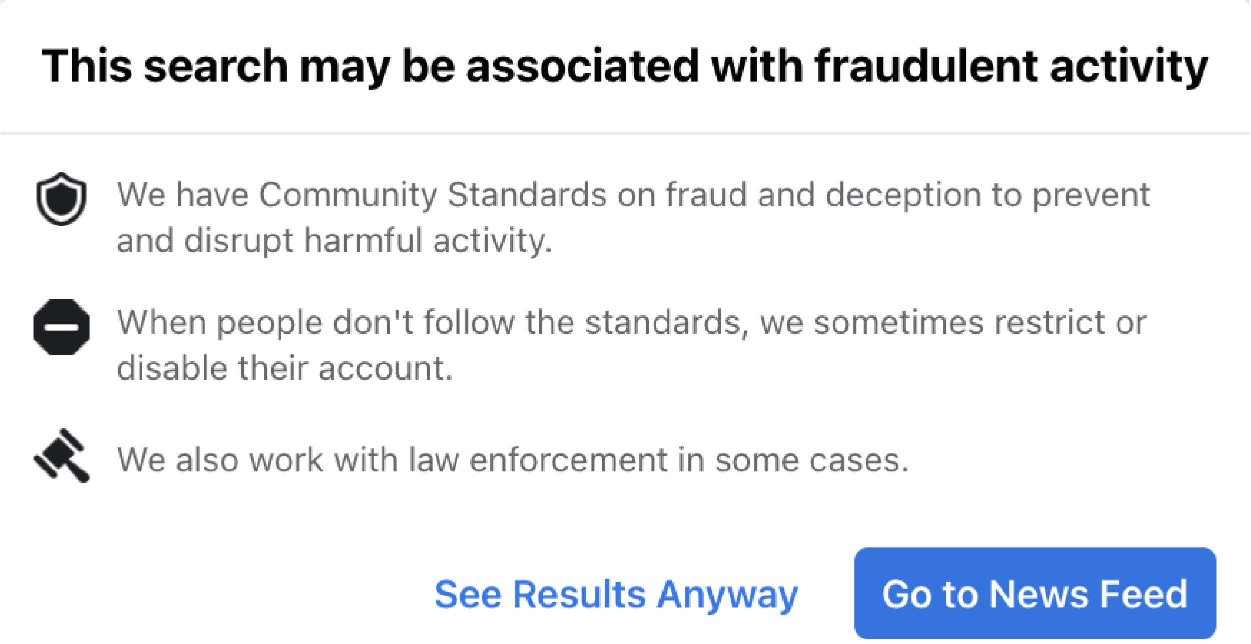
We also use a mixture of nudge behaviour and proactive warnings via messenger to let users know when they are messaging an account which is demonstrating behaviour similar to that we have previously seen from scammers. These accounts have not breached the levels we would need to see to suspend or disable an account but are suspicious enough to warn users about.

On Facebook we also deploy a search intercept which gives users a warning when they search for a fraud term e.g. "carding" (the trafficking and unauthorised use of credit cards). The warning will tell them that the term they are searching for might be associated with fraudulent activity, and warn them we might restrict their activity if they violate our policies.



Are you sure you want to continue?

The term that you've searched for is sometimes associated with fraudulent activity, which isn't allowed on Facebook.

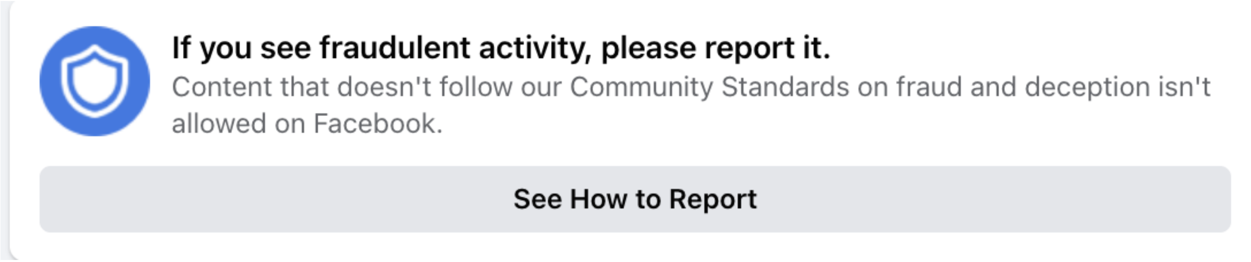


This search may be associated with fraudulent activity

We have Community Standards on fraud and deception to prevent and disrupt harmful activity.

When people don't follow the standards, we sometimes restrict or disable their account.

We also work with law enforcement in some cases.



If you see fraudulent activity, please report it.

Content that doesn't follow our Community Standards on fraud and deception isn't allowed on Facebook.

Whilst our aim is to catch content proactively, ideally before users report it to us or interact with it, where users do come across such content we want to make the process of reporting it to us and getting it taken down as easy as possible.

Our in app reporting function is available via the “three dots” that appear in every piece of posted content and users can report organic content which they consider to be harmful in some way or advertising content which they no longer want to see or think is irrelevant or inappropriate. These reports are an integral part of training our systems to better spot fraudulent activity.

Reports of Fraudulent Ads also come from third parties like the Financial Conduct Authority (FCA) who are onboarded to our Consumer Policy Channel (CPC). The CPC enables us to work with consumer protection bodies, Government departments, regulators and law enforcement to help us better detect and remove content that violates our policies or local law by taking action on content reported to us by agencies who have the appropriate authority to make determinations in relation to the commercial content or activity they are reporting. We generally deal with reports of commercial content and activity received via our Consumer Policy Channel within 24-48 hours.

Where we see a trend towards a particular type of activity that is not captured by our policies we review those policies with the input of experts to ensure they remain fit for purpose as the landscape evolves.

Fraud perpetrated through paid-content/advertising

For fraudulent activity using advertising on our platforms we also focus on behaviours rather than content, given the ever changing nature of these scams. These efforts are geared towards building more proactive tools to automatically take down this content before it goes live using a combination of AI and human review.

Our systems incorporate signals such as user feedback, fake/compromised account signals and ad content signals and tactics which go into building our proactive detection technology. We’ve also invested in ensuring our specialised reviewers can understand and identify this content - which by its very nature is hard to spot. Relative to other harms on Facebook, the scams space is more complex and difficult for reviewers to accurately classify, so we have sought to build a more holistic understanding of the abuse over time.

Enforcement

When we find ads that violate our policies, we may go beyond simply rejecting the ad - we disable ad accounts and remove their ability to advertise in the future. We may also disable the responsible business manager and prevent the user profile and the page associated with the abuse from further advertising on our platform.

We're also leveraging ways to identify bad actors more quickly and prevent policy-violating ads before they happen. For example, when we have signals of a suspicious advertiser or we’ve seen a history of policy violations, we take action, like limiting or disabling their ad activity.

In some cases, these bad ads come from compromised or fake accounts. We shared in our last enforcement report that we have removed billions of fake accounts from the platform.

Ad Review Process overview

Ads on Facebook and Instagram are not only subject to our Ad policies, they are also subject to our **Ad Review System**. The system reviews different elements of an ad - like text, and images - for certain violations. If we detect a violation of our Ad policies, we will reject the ad. This review happens before ads become visible to people but may also happen after, if people hide, block, or provide negative feedback about an ad and we therefore have reason to believe there may be a problem. We use this kind of feedback and reporting to not only take action on the ad in question but also to help train our automated systems to find and flag problematic ads more effectively over time. In the case of people who repeatedly violate these policies, we may remove their Account or Page.

We have also made all active ads available to see in our Ad Library. People can report the ads from within the library, or learn more about the ad. This critical transparency measure enables the public, including advocates and regulators, to better understand what organisations are running ads and who is seeing them.

The challenge with detecting fraud is that the fraudulent ad or piece of content is specifically designed to appear to be legitimate both to our systems and to the user. Once it has been reported to us we are able to feed the learnings from each report into our systems in order to improve detection and respond to new tactics by the scammers. However, this is a very adversarial space with scammers using increasingly sophisticated methods to avoid detection.

Stakeholder collaboration

Our priority is always to act against a bad actor as quickly as possible for any violation, but we are operating in a particularly adversarial space with bad actors who use increasingly sophisticated means to avoid detection. It is therefore vital for us to work with and learn from external stakeholders on their approach and to share expertise; this is a complex issue that requires a joined up multi-stakeholder approach.

We are currently partnering with the **Advertising Standards Authority** on its Scam Ad Alert System, working with over 20 ad networks and advertising platforms to act on reports of potential scam advertising. Consumers can report scam ads appearing in paid-for space online via a form on the ASA website. The ASA then sends an alert to all participating platforms with key details of the scam ad. The participating platforms will then remove the offending ad and suspend the advertiser's account, or add them to a blocklist (if the ad has not already appeared on the platform) in order to prevent the ad from appearing in the future. Between 1 March 2021 to 25 March 2022 the UK Scam Ad Alert system has received 1,251 reports from the public; which has resulted in 67 alerts being sent to online platforms. 14 (21%) related to ads seen on social media sites.

We also leverage reports from other partners to help us identify policy-violating content. These reports allow us to respond more quickly in what is often a dynamic and complex environment. In this regard, we work closely with a number of UK regulators including **the FCA, CMA and National Trading Standards** to help us better detect and take action on content that is either in breach of our policies or in breach of local laws. We have also worked with the FCA on its Scamsmart campaign to help it raise awareness of investment scams.

Where there is sufficient evidence and a criminal element, like credit card or payment fraud, we can engage with UK law enforcement through our global Law Enforcement Outreach Team; collaborating with **the City of London Police**, as the national lead police force for fraud, and **The Dedicated Card and Payment Crime Unit (DCPCU)** - a unique proactive police unit, with a national remit, formed as a partnership between **UK Finance, the City of London Police** and **the Metropolitan Police** together with **the Home Office**.

Through this successful partnership, Police provide us with information about various types of scams they have identified, together with users involved and violating content, enabling us to react quickly to remove it. We are also in discussion with Police and Crime Commissioners about our general approach in this area.

In April 2021, **the Online Fraud Steering Group (OFSG)** was set up, co-chaired by **techUK, UK Finance and the National Economic Crime Centre**, to form collective solutions to respond to patterns of fraudulent activity. Since its inception, the OFSG, the Online Fraud Delivery Group and its sub-groups have met at least thirty times; representatives from Meta's UK Public Policy team attend. Other attendees include other tech companies (**Tik Tok, Google, Microsoft, Snap, Twitter and Amazon**); representatives from the banking industry; law enforcement agencies; Government (**Department for Work and Pensions, Department for Digital Culture, Media and Sport, and the Home Office**) and the **FCA**.

In a short time, the group has agreed a delivery infrastructure, operational principles, and governance. Supporting the Home Office's upcoming 2022 - 2025 Fraud Action Plan the group aims to:

- render the UK the least attractive place for online fraudsters to operate;
- involve all relevant sectors as required to collaborate and form targeted responses to prevent different types of fraud;
- share information and best practices to ensure a shared understanding around online fraud and its complexities;
- bring improved coordination between law enforcement and the tech and banking sectors; and
- enhance public communication around the complexities of financial fraud and promote consumer awareness.

Four key workstreams have begun work to cut across different fraud typologies: addressing fraud through online advertising; developing a threat assessment; enhancing communications and education; and, striving for innovative and preventative solutions.

As part of our work with the OFSG in December 2021 Meta committed to introducing a revised advertising onboarding process that requires UK regulated financial services to be authorised by the Financial Conduct Authority prior to serving financial services adverts on our sites.

Late last year we also began requiring advertisers targeting people in the UK (and across Europe) to self-declare if their ads pertain to housing, employment or credit. These advertisers must submit to significant advertising restrictions and additional transparency in the Ad Library. If an advertiser promotes loan services, for example, but fails to make this declaration we will reject its ad(s).

While this does not cover the full breadth of the financial services sector, it reflects our commitment to protecting consumers from bad actors engaged in predatory or discriminatory lending practices.

As well as tackling violations of our policies on the platform, we're also focused on prevention and giving direct support to people impacted by scams. We've donated £3 million to **Citizens Advice** to help them set up a UK anti-online scams initiative for this purpose, which includes a telephone helpline for people who have been scammed online and face-to-face consultations for serious cases.

Increasing consumer awareness about types of fraudulent activity online is an important part of this collaboration. On 15 September 2021, tech companies, including Meta pledged to support **Take Five to Stop Fraud**, the anti-fraud campaign run by **UK Finance**. The technology companies collectively donated $1 million worth of advertising to the campaign which will help publicise the Take 5 to Stop Fraud advice to consumers and enable it to reach a significant proportion of the online population with these messages.

Most recently in March 2022 Meta joined **Stop Scams UK**, an industry-led not for profit collaboration of responsible businesses from across the banking, technology and telecoms sectors that have come together to stop scams at source. Stop Scams UK was set up with support from Ofcom and the FCA and we believe it can make a real difference by bringing together knowledge, insight and expertise from these three core sectors.

We hope this demonstrates our commitment to tackling fraud and scams on our platform and our desire to work with stakeholders. We strive to make our platforms safer for people and businesses, by continuously working to improve our detection and enforcement practices to stay on top of ever-evolving malicious threats.

22 April 2022