Recommendations for Ofcom's Guidance under section 65 of the Online Safety Act, regarding the user identity verification duty

Introduction

This submission from Clean Up The Internet is our response to Ofcom's call for evidence for the Third phase of online safety regulation. It concerns section 64 of the Online Safety Act, which requires Category 1 platforms to offer all adult users an option to verify their identity, and section 15(9), which requires the same platforms to allow users to filter out non-verified users. It sets out 5 recommendations for the content of the Guidance which Ofcom is required to produce under section 65.

The s64 user identity verification duty, and associated s15(9) filter, provide an opportunity to reduce the role played by anonymous and fake accounts in a very wide range of online harms on Category 1 Platforms, through giving users more choice and more control. This includes some of the most serious illegal harms designated "priority offences" under the OSA. It also includes content which may not meet a criminal threshold, but is experienced by users as abusive, or unpleasant, or misleading and which cumulatively has a significant negative impact on their online experiences.

The link between anonymous and fake accounts and online harms is widely understood, as is the potential for user verification to help reduce these risks. Numerous pieces of public attitude research and opinion polling have found that a large majority of UK social media users see anonymous and fake accounts as a significant factor in a range of harms; are keen for more options to identify and avoid non-verified accounts; and would be willing to verify their identity[1][2].

The user identity verification measures therefore have the potential not only to make a considerable contribution to tackling a very wide range of online harms, but also to be

1 Opinium for Compassion in Politics, 2022: 81% say that they would be willing to provide a piece of personal identification to a social media platform in order to receive a "verified" account; 72% would choose to remove all unverified user-content from their feed if that option was available. https://www.compassioninpolitics.com/4_in_5_back_efforts_to_curb_toxic_anonymous_social_media_accounts

2 Opinium for Clean Up The Internet, 2023: 78% of UK social media users believe that it would be helpful to be able to see which social media accounts have been verified, to help them avoid scams. https://www.cleanuptheinternet.org.uk/post/new-research-78-of-social-media-users-say-being-able-to-see-who-s-verified-would-help-avoid-scams

among the best understood, most visible, and most popular changes to stem from the OSA.

For this considerable potential to be realised, Ofcom's guidance will need to set the right framework for platforms to follow. In the absence of regulatory supervision, platforms have an extremely poor record of mitigating harms from fake and anonymous accounts, or of implementing robust and effective verification schemes. Without sufficient clear and robust guidance, there's every reason to fear more of the same delay, obfuscation, platitudes and half-measures which have characterised platforms' action in this area to date.

This paper therefore sets out five recommendations for Ofcom as it develops its Guidance:

1. The guidance must address the full range of relevant harms where fake and anonymous accounts are a risk factor

2. The guidance must seek to raise the bar, and must not simply collate existing "best practice" which sadly is better described as "worst practice".

3. The guidance should align with other relevant OSA codes, and other relevant regulation, standards, and guidance

4. The guidance must set out criteria which platforms must satisfy for an identity verification scheme to fulfill the duty.

5. The guidance should offer examples of methods and processes which will not be considered to satisfy the duty, and some which could be.

Recommendation 1: The guidance must address the full range of relevant harms where fake and anonymous accounts are a risk factor

The wording of sections 64, 65, and 15(9) is quite brief. Ministers repeatedly assured parliamentarians during the legislative process that the expectation was not for Ofcom's guidance to be similarly brief, or to lack rigour or specificity. On the contrary, brevity in the primary legislation was intended to give Ofcom sufficient flexibility to devise comprehensive guidance that ensured platforms developed effective new verification schemes, and to update it over time in response to new evidence or new technology.[3]

3 For example at Commons Committee Stage, responding to a proposed amendment to expand the wording of what became section 65, then minister Chris Philp justified brevity by explaining "we want to ensure that there is flexibility for Ofcom, in writing those guidelines", and that "We think that the powers set out in clause 58 [which became section 65] give Ofcom the ability to set the relevant regulatory guidance"

Ofcom staff have observed to us in meetings that one aspect of the brevity of these sections is that they do not state a specific purpose for the verification duty, or a specific harm that verification or the filter is intended to address. We would agree. In the absence of any limitations or qualifications in the specific sections, the most logical interpretation of sections 64, 65, and 15(9) is that they are intended to serve the overall purpose of the Act, as set out in Section 1, of "making the use of internet services regulated by this Act safer for individuals in the United Kingdom". It was clearly the expectation of the government, and legislators, when announcing the measure and voting through these clauses, that they would empower users to reduce their risk of exposure to the full range of harms which can be associated with fake and anonymous accounts.

In Volume 2 of Ofcom's recent consultation its "Protecting people from illegal harms online" fake and anonymous accounts are identified as a risk factor for 15 different priority offences. Indeed, fake and anonymous accounts are highlighted as functionalities which "stand out as posing particular risks". Ofcom then notes in Volume 4 of the same consultation, when considering a mandatory verification scheme, that "given the broad range of illegal harms of which anonymity increases the risk, any interference could be said to be in pursuit of several aims, including the prevention of disorder or crime, the protection of the rights and freedoms of others, and the interests of national security." It further notes, in its consideration of notability or monetised schemes, that poorly designed or insufficiently robust schemes which purport to offer verification can themselves be a risk factor for illegal harms, by making inauthentic and deceptive accounts more plausible.

As well as illegal harms, the user identity verification duty and user empowerment duties are envisaged as also providing users with more choice and control over content which may not reach a criminal threshold, but which users may nonetheless find very upsetting, offensive, or harmful. They are envisaged as playing this broader role – sometimes referred to in the legislative process as part of a "triple shield" – because, by relying on offering users choices and optional controls – they entail fewer trade-offs around freedom of expression than other measures such as content removal, or user bans.

Crucially, there is nothing in the fact that verification is envisaged as providing users options to protect themselves from legal content which conflicts with the contribution these duties should also make to protecting users from illegal harms. Ofcom's Guidance must therefore ensure that the user identity verification duty, and accompanying filter, address the full range of harms for which anonymous and fake accounts are a risk factor, and for which verification measures can therefore act as an essential mitigation.

This must include giving full consideration to risks of illegal harm, as identified in Ofcom's own Register of Risks for illegal content. It must also include consideration of harmful,

---

https://hansard.parliament.uk/commons/2022-06-07/debates/90e5ab5b-a47b-4750-9c47-3f0ac7cd46eb/OnlineSafetyBill(SixthSitting)#

upsetting, and offensive content which may not meet a criminal threshold, including the cumulative impact of such content on vulnerable individuals. When Ofcom considers the proportionality of including a potential recommendation in its Guidance, it must consider the impact and costs – to individuals, society, and businesses – of all the harms for which anonymous and fake accounts are a risk factor.

Recommendation 2: the guidance must seek to raise the bar, and must not simply collate existing "best practice".

In the absence of regulatory oversight, platforms' approaches to account creation and identity verification have consistently failed to make any significant dent in the problems associated with anonymous and fake accounts. In addition there has been significant obfuscation on the part of the platforms as to both the scale of the problem and the effectiveness of their current approaches in tackling it - for example Twitter's dishonest claims about the limited role of anonymous accounts in abuse of England footballers during the Euro 2020 tournament, which appeared to rely on a definition of "not anonymous" which would include, literally, Mickey Mouse accounts.[4]

The government, and parliamentarians, recognised this failure on the part of platforms. They introduced specific new duties to ensure users gained new options to verify their identity, and to identify and avoid anonymous accounts, because they wanted to introduce new benefits and new protections for users.

When announcing the introduction of the measures into the Bill, then Secretary of State Nadine Dorries talked of "*new measures* to put *greater power* in the hands of social media users themselves". The same press release noted that at present "the vast majority of social networks used in the UK do not require people to share any personal details about themselves - they are able to identify themselves by a nickname, alias or other term not linked to a legal identity." In other words, there was a clear expectation that the introduction of these new measures would deliver something concrete, and would entail a significant change from the status quo for "the vast majority of social networks".

All of this means there is very limited evidence of "best practice" for Ofcom to draw on. The legislative intent is clearly to improve practice, not to consolidate it. So whilst it's perfectly reasonable for Ofcom to request evidence relating to actually existing verification schemes, it must not be constrained by them.

4 https://www.cleanuptheinternet.org.uk/post/twitter-s-anonymity-claims-appear-to-rely-on-classifying-mickey-mouse-accounts-as-not-anonymous

Recommendation 3: the guidance should align with other relevant OSA codes, and other relevant regulation, standards, and guidance.

For users, trust and usability of verification schemes will be greatly enhanced if there is a reasonable degree of similarity and consistency in how they operate. Users should be able to expect similar standards and principles to apply across different verification processes which they encounter. “Verification” should, in different contexts, still signify for example a similar level of rigour, a similar level of real world reliability, a similar level of trustworthiness. Consistency will increase users’ familiarity, leading to wider take-up and more informed use of both verification and the filter on non-verified accounts.

Platforms, and third party providers of verification services, will also benefit in the longer term from consistency between what Ofcom requires in its guidance for the user verification duty, and requirements for other contexts and applications for the user identity verification duty. It would ensure greater efficiency and lower costs, and open up more potential for innovation.

This means Ofcom’s guidance should identify, and as far as possible align with, other relevant standards and good practice. This should include the age assurance guidance under the OSA, and other UK government, and international standards and good practice guides for identity verification, such as the Cabinet Office Good Practice Guide 45, the Trust and Identity Framework, and the EU's eIDas framework. Ofcom should indicate to platforms standards which, if met, would ensure that their verification methods and processes comply with their obligations under the user identity verification duty.

Recommendation 4: the guidance must set out criteria which platforms must satisfy for an identity verification scheme to fulfil the duty.

As detailed above, the user identity verification duty is intended to change platform behaviour and raise standards. The OSA reflects a recognition that, left to their own devices, few platforms have failed to develop satisfactory approaches to user verification, or tackling the harms associated with anonymous or fake accounts. It is therefore imperative that Ofcom ensures minimum standards are met, by setting clear criteria which platforms’ methods and processes must meet.

We suggest that these criteria should include:

a. **Accuracy, robustness and reliability**: For the user identity verification duty and filter to genuinely empower users to reduce their exposure to the harms associated with anonymous and fake accounts, platforms' verification methods and processes need to be sufficiently accurate, robust, and reliable. Furthermore, if they are not, then flawed verification processes could actually exacerbate some of the problems associated with fake accounts - for example if scammers are able to easily gain "verified" status for fake accounts. Ofcom's guidance should therefore set stringent minimum standards for how effective a platform's methods and processes are at accurately determining a user's identity, preventing circumvention from bad actors, and delivering reliable and repeatable results which other users can trust over time. It should include requiring platforms to set, state, and report against their own targets for accuracy, robustness and reliability. It should set an expectation that the effectiveness of verification processes is regularly reviewed, with a view to quickly identifying and closing any loopholes or responding to new circumvention tactics by bad actors. It could set out thresholds or triggers for enhanced verification checks, for example if any account exceeds a certain level of reach or activity, or engages in certain types of activity or content, or a particular method of verification is found over time to have become vulnerable to circumvention.

b. **Accessibility:** S64 states that plaforms must offer verification to "all adult users", and s65 further states that Ofcom's guidance should have "particular regard" to the desirability of verification being available to "vulnerable adult users". Ofcom's guidance therefore should set out some clear steps which platforms are expected to follow, and levels of accessibility which they are expected to meet, to ensure the choice to verify is genuinely widely available. This should include both on-paper assessments of how accessible their processes and methods are likely to be (for example - if documentation is required, what proportion of the population has access to an acceptable form of documentation) and monitoring of real-world access rates by different groups, including by different protected characteristics. Ofcom should include recommendations for verification methods and processes which could improve accessibility for vulnerable groups - for example "vouching" to cater for individuals who lack access to documents. Ofcom should require platforms to make clear to all users what tools and support are available to help them access verification.

c. **Affordability:** Ofcom should make it clear that, in order to be accessible to all adult users, verification should be made available at no extra cost to users. It is crucial that this is made clear because some platforms have recently sought to monetise verification as part of "premium" subscriptions. The Online Safety Act envisages that verification be made available to all users of Category One services, i.e. should be a core safety feature. Platforms should not be able to place core safety features behind a paywall, any more than car manufacturers are allowed to charge extra for brakes or airbags.

d. **Visibility of verification status:** For verification to deliver on its potential to reduce harms associated with deception and/or impersonation, users need to be able to see whether or not another user is verified. There is considerable evidence of user demand for verification to be visible. Ofcom's own research into measures users would find helpful to avoid fraud and scams, found that a 'warning from the platform that content or messages come from an unverified source' is the most popular measure platforms could introduce.[5] Polling commissioned by Clean Up The Internet found that 78% of UK social media users believe that it would be helpful to be able to see which social media accounts have been verified, to help them avoid scams. Almost as many also said being able to see which accounts have been verified would help with identifying bullies or trolls (77%); spotting false or misleading news stories (72%); and buying products or services (68%).[6] Ofcom should stipulate that verification status should be visible, and offer guidance as to best practice for labeling to ensure a degree of consistency across different platforms. Ofcom should require platforms to take reasonable steps to prevent the mimicry of "verified status" symbols to confuse users – for example, if a tick is used to signify verified status, non-verified users should not be able to use images of ticks in their profile photos, or tick emojis in their usernames.

e. **Account security and user authentication**: A foreseeable potential consequence of the user identity verification being effective, and reducing the ability of non-verified accounts to deceive and/or harm UK users, would be increased efforts by bad actors to hack and hi-jack verified accounts. Ofcom should therefore set expectations that users of verified accounts are offered, educated about, and nudged towards, security features which mitigate this risk, such as for example two factor authentication. Ofcom should outlaw bundling such core security features with paywalled "premium" subscriptions.

f. **Privacy and data security**: For users to trust platforms' verification systems, they will need to trust the privacy and security of the processes and systems. Ofcom should make it clear that it expects platforms to comply fully with all relevant data protection rules including UK DPA and GDPR and the Age Appropriate Design Code.

g. **Encouraging user awareness and uptake**: For the user identity verification duty, and accompanying filters, to work as legislators intended, a critical mass of UK users will need to make use of them. There is strong evidence of public appetite for these features. However there is a risk, given platforms' resistance to such measures to date, that some may be tempted to "hide" or "bury" the features to

5 https://www.ofcom.org.uk/research-and-data/online-research/online-fraud-and-scams

6 https://www.cleanuptheinternet.org.uk/post/new-research-78-of-social-media-users-say-being-able-to-see-who-s-verified-would-help-avoid-scams

suppress uptake. Ofcom's guidance should therefore set a clear expectation that platforms prompt users about the new features when they become available, and at regular intervals thereafter. Ofcom should stipulate that both the verification option, and the filter option, are easy to find and use from app and website homescreens. Ofcom should set minimum penetration levels, and should require platforms to state, and report against, their own targets for levels of user awareness and user uptake of both verification and the filter on non-verified accounts.

h. **Interoperability and user choice -** Platforms accepting a variety of different methods of verification, including those provided by third parties (such as providers of digital wallets, for example), would have considerable potential benefits for users. Choice in how to verify would promote accessibility, by increasing the likelihood of a user finding an option which works for them. Accepting third party options would reduce the burden on users, by reducing the time and information required to verify on each new platform they wish to access, if they have already verified their identity to an accepted standard elsewhere. It would mean users were able to choose to go through the verification process with a company or platform they trusted - and then have options to minimise the information they needed to share with other platforms. Ofcom should state that it expects platforms to offer users choices of how to verify, including accepting verification via third party providers. The guidance should promote interoperability by setting out common standards, technical frameworks and other specifications for verification systems to follow.

Recommendation 5: the guidance should offer examples of methods and processes which will not be considered to satisfy the duty, and some which could be.

Alongside setting out criteria for platforms to meet, Ofcom should offer some clear, worked examples of approaches that are unlikely to satisfy the duty, and some which, assuming they are implemented properly, are more likely to be compliant. This would offer more clarity, sooner, for platforms, potential third party verification providers, and users.

Many platforms have a track record of obfuscation around verification and anonymous accounts in the past. Ofcom should act to preempt this by making clear that certain features which platforms have described as "verification" will not on their own satisfy the duty, including sending a confirmation link by email or SMS, and use of self-declaration checkboxes.

Conclusion

UK internet users have an understanding that fake and anonymous accounts carry higher levels of risk, and have an appetite for the verification and filter measures envisaged in the OSA as ways of managing those risks. If the measures are implemented effectively, they could become one of the most effective, and most popular, aspects of the new regulatory regime.

However, this to is unlikely to happen unless Ofcom's guidance sets the right framework. Managing risks of anonymous and fake accounts, and offering users robust identity verification options, are areas which platforms have a long track record of failure. We can only expect change to come if the regulator requires it.

Given the considerable public appetite for the measures, and their considerable potential to reduce harm, Ofcom should seek to ensure that that progress from here is as swift as possible. Crucially, the first version of its guidance needs to aim to be sufficiently stringent to deliver genuine benefits. Setting a very low bar in its first version of the guidance, on the grounds that it can subsequently "iterate up" if it needs to, is unlikely to deliver the changes which the Act requires and users expect. Not only could it delay the benefits which the verification measures can deliver, but even worse if users' first experiences of the OSA's verification measures are ineffective or unreliable, it could erode public trust and appetite for such measures for the long term.

We appreciate the opportunity to offer input at this earlier stage of the thinking. We hope that Ofcom finds it useful in preparing its draft guidance, and look forward to being consulted further on that draft. We would welcome the opportunity to discuss these measures. We suggest that Ofcom might find it helpful to convene a roundtable with all relevant stakeholders, including platforms, to work through the issues as speedily as possible.

Clean Up The Internet, May 2024