How requiring VLOPs to offer their users options to verify their identity could contribute to protecting election integrity

1. Introduction

Clean Up The Internet is an independent, UK and EU-based, not-for-profit organisation concerned about the degradation in online discourse and its implications for democracy. Our EU transparency register ID number is 099046239388-80.

We campaign for evidence-based action to increase civility and respect online, to safeguard freedom of expression, and to reduce online bullying, trolling, intimidation, fraud and disinformation.

We are delighted to have the opportunity to submit our views on the draft DSA guidelines for election integrity.

We agree with the Commission’s current proposals, but have an additional, complementary suggestion for a specific mitigation measure which the Commission could add to its 3.2.c “Measures to provide users with more contextual information on the content and accounts they engage with”.

We propose, as an additional measure for any VLOP which offers users the ability to create accounts and profiles, that they implement an optional user identity verification scheme which:

- offers all their users options to voluntarily verify their identity
- clearly labels which accounts/profiles are and are not verified, so that the verification status of each user is visible to all users of the platform
- offers their users controls to manage their interaction with non-verified accounts, up to and including an option to block all interaction from non-verified accounts

Such a scheme would help address the systemic risks posed by fake accounts, which are widely recognised as a key tool for disinformation and influence operations. It would also provide users with additional options to protect themselves from the wider range of harms associated with fake and anonymous accounts.

Clean Up The Internet has made similar proposals within the United Kingdom’s Online Safety Act process. Under that legislation the largest platforms (“Category
One”) will in due course be subject to a “user identity verification duty”, which will require them to offer optional verification, combined with optional filters on non-verified accounts, similar to the measures we are proposing here. There’s therefore an opportunity to align these measures.

2. How fake social media accounts enable electoral interference

Fake accounts are a crucial enabler of electoral interference on social media. This is acknowledged by the platforms - Meta says of co-ordinated inauthentic behaviour that “fake accounts are central to the operation”.

The ability to create fake accounts enables influence operations to create networks of accounts under their control (which may be automated, partially automated, or non automated) which can:

- post inauthentic content
- deploy inauthentic reactions (e.g. likes or comments)
- inflate numbers of followers or subscribers
- create inauthentic groups, pages or fora

Such inauthentic activity serves both to deceive other genuine users, and to game recommender algorithms to secure greater amplification of inauthentic content. Platforms deploy, to varying degrees, efforts to detect and remove fake accounts. However, these efforts are by their very nature post hoc, and whilst they occasionally disrupt specific operations or networks, they have not prevented fake accounts continuing to be central to the core tactics, techniques and procedures upon which foreign influence operations rely.

To offer just a few very recent examples:

- In September 2022 Meta reported having taken action against a “Russian network [that] targeted primarily Germany, France, Italy, Ukraine and the UK, with narratives focused on the war in Ukraine and its impact in Europe. The largest and most complex Russian operation we’ve disrupted since the war in Ukraine began, it ran a sprawling network of over 60 websites” - but only after the network had first been documented by German investigative journalists at public service broadcaster ZDF.

- In its report under the DSA Disinformation Code, from March to June 2023, TikTok disclosed having detected and removed 5,885,958 fake accounts purporting to be within the EU, which had accumulated 47,409,587 followers at time of detection and removal. This included 9,246 accounts, which had accumulated 356,935 followers, purporting to be in Slovakia - in the run-up to crucial elections. In the same period, Meta reported that “fake accounts
represented approximately 4-5% of our worldwide monthly active users (MAU) on Facebook”.

- In July 2023, inauthentic accounts were found to have played a key role in influencing online political discourse around the Uxbridge and South Ruislip by-election in the United Kingdom. Inauthentic accounts greatly amplified hostility to the London Mayor’s “Ultra Low Emission Zone”. Not only did this potentially influence the result, a narrow surprise win for the Conservative Party, but it also appeared to influence the direction of government policy, with the Uxbridge and South Ruislip result linked to subsequent changes in the tone and content of UK government’s climate policies.

- In September 2023, Microsoft reported having uncovered Chinese social media influence operations that “deploy thousands of inauthentic accounts across dozens of websites, spreading memes, videos, and messages in multiple languages”. It warned that “ahead of the 2022 US midterms, Microsoft and industry partners observed CCP-affiliated social media accounts impersonating US voters—new territory for CCP-affiliated IO [influence operations]. These accounts posed as Americans across the political spectrum and responded to comments from authentic users.” The report noted that “unlike earlier IO campaigns from CCP-affiliated actors that used easy-to-spot computer generated handles, display names and profile pictures, these more sophisticated accounts are operated by real people who employ fictitious or stolen identities to conceal the accounts’ affiliation with the CCP”.

It’s worth noting that fake accounts are also used to commit many other crimes and perpetrate many other forms of harm on other users. The UK’s regulator, Ofcom, identified in its draft Register Of Risks related to illegal content, that fake accounts were a risk factor for all of the following: Grooming; CSAM; Suicide and Self-Harm; Harassment, stalking, threats and abuse; Hate offences; Drugs offences; Firearms offences; Extreme pornography; Intimate www.cleanuptheinternet.org.uk 2 [?] image abuse; Fraud; Foreign interference offence; False communications offence; Cyberflashing. The techniques, tactics and procedures deployed by foreign influence operations often overlap with those used by other criminals.

Valent has highlighted the similarities between political disinformation efforts including those by foreign state actors, and the “coordinated influence campaigns [that] are a major feature of online discussion about crypto and NFTs.” The NATO Strategic Communications Centre of Excellence found that frequently the same networks of inauthentic accounts may be deployed for both political and “commercial” manipulation of social media conversations, finding that “While we did identify political manipulation—as many as four out of five accounts used for manipulation on Facebook had been used to engage with political content to some extent—we assess that more than 90% of purchased engagements on social media are used for commercial purposes.”
3. How an optional verification scheme would help disrupt election interference operations

We suggest that the Commission recommends, within its guidelines, that VLOPs implement a measure to:

1. **Give all social media users the right to verify their identity if they choose.** Every social media user should be given the option of a robust, secure means of verifying that the identity they are using on social media is authentic. Users who wish to continue unverified should be free to continue to do so.

2. **Ensure clear and visible indications to make it easy for everyone to see whether or not a user is verified.** Each user would then be able to bring their own judgement as to what a verification status might say about the credibility and reliability of another user’s content.

3. **Give users the option to block interaction with unverified users** Some users will have more appetite to hear from, and interact with, unverified users whereas others will have a lower appetite for risk. This should be a matter of individual user choice. Every social media user should be offered options to manage their level of interaction with unverified users, including an option to block communication, comments and other interaction from all unverified users, as a category and pre-emptively.

We suggest that such a measure would fit very naturally into section 3.2.c, “Measures to provide users with more contextual information on the content and accounts they engage with.”

Such a measure would not prevent Influence Operations from creating fake accounts. However, it would significantly reduce the potency of these accounts and their ability to manipulate discourse:

- All users would be provided with an absolutely crucial piece of information to assess the trustworthiness of a source, strengthening all the other “Measures to provide users with more contextual information” suggested in section 3.2.c. Where a fact or an opinion was being offered and endorsed by accounts which had chosen not to be verified, users would be empowered to bring their own judgement as to what this might mean for the reliability of the information or the authenticity of the opinion.

- In addition, users would have the option of avoiding non-verified accounts entirely. This would mean users who were concerned about disinformation, or other crimes associated with fake accounts, had a very easy-to-use tool to protect themselves.

- Platforms’ recommender algorithms would be less vulnerable to manipulation from fake engagement, because they would be able to distinguish between engagement by verified and non-verified accounts, and take steps to investigate any suspicious discrepancies in engagement from these two categories.
4. How an optional verification scheme could work

It may not be appropriate for the Commission to stipulate the precise approach to verification. Different VLOPs could develop methods and processes which are appropriate to them and their users. The Commission should simply set some minimum standards which require VLOPs to demonstrate that their process is sufficiently robust to impede bad actors.

Several different kinds of identity verification would be available to platforms, including:

- **Verification by checking a user’s claimed identity against some form of photographic ID.** This method is already used by a range of platforms in a range of circumstances - for example Meta’s monetised premium subscription, “Meta Verified”, requires this. For those EU member states which do not issue a universal identity card, this would need to include other documents such as passports and driving licenses. To be robust, such verification systems need to be able to check both the ownership and the authenticity of the ID, which may involve additional processes such as requiring a user to upload a video selfie, and checking the provided document against lists of lost/stolen IDs.

- **Verification by checking a user’s claimed identity against digital identity credentials already held by a trusted third party.** This method would enable users verify their identity to a platform by sharing credentials from a reusable digital ID which they have created and verified with a third party. This could be either a private sector digital identity provider, such as Yoti, or a digital identity provided to citizens by member states. This method could provide an important additional use case for the planned EU Digital Identity Wallet.

- **Verification by seeking confirmation of a user’s claimed identity by other trusted individuals, often known as “vouching”.** This method is commonly used in more “analogue” situations - for example signing the back of a passport photo to verify the true likeness of someone - and is also commonly used to include people whose vulnerabilities make them less likely to be able to afford or access other forms of verification (e.g. people with no permanent address). In some use cases, this method requires the individual providing the vouch to have some form of professional status e.g teacher, lawyer or doctor. Alternatively, or additionally, the provider of the vouch could be required to themselves be a verified user, with a certain period of stable use of the platform. Further safeguards against abuse could include limits on the number of vouches any one user/individual can provide, or potential consequences for the user providing the vouch if the vouched-for user violates the platform’s terms of service. This could be offered as a “last resort” method for vulnerable users for whom other methods are not accessible.

A related consideration would be whether the verification process is conducted by the platform itself, or by a third party provider, and therefore who handles and/or
retains any relevant user data. There are examples of “monetised” and “notable” schemes that take each approach. For example “Meta Verified”, which requires users to provide government ID documents, appears to be conducted in-house, whereas X Premium, not currently available in the EU, states that it uses a third party provider.

Whilst all relevant privacy and data protection rules should apply whoever administers the scheme, some users may feel more comfortable sharing their data with a trusted third party, or others may feel more comfortable sharing their data directly with a platform with which they are already familiar. Offering users choices, including choices of both mechanism and provider, would help maximise accessibility and therefore take-up.

Another related consideration will be account security. Given that a voluntary verification scheme would make anonymous and fake accounts less potent tools for criminality, we could reasonably anticipate bad actors increasing efforts to hack and steal verified accounts. Platforms could combine the verification process with requirements that users implement security measures such as stronger passwords and two-factor authentication. Reporting mechanisms for hacked and stolen verified accounts would need to be sufficiently resourced and responsive.

5. Rights considerations

Any potential restriction on the ability to use social media anonymously should be considered carefully from the point of view of its impact on users’ fundamental rights. Such a consideration needs to account for the fact that anonymity can be crucial to safeguarding some users’ rights, such as to privacy or to freedom of expression - for example a whistleblower, or a vulnerable minority. However, anonymous and fake accounts can also be enablers of hate speech and disinformation which often disproportionately threatens those very same minorities. In addition, any verification scheme has the potential to present particular barriers to vulnerable groups who may be less likely to have access to, for example, a government-issued identity document, thus impinging on their freedom of expression.

An optional verification measure, which requires only that users are offered a choice as to whether they verify, hugely reduces any negative impact on a user’s privacy rights, because users can choose whether or not to go through the process. Any remaining impact would be mitigated further by ensuring that services have in place robust measures to comply with their data protection obligations, and by giving users choices as to how they verify, including choices to verify via third parties. Users could then choose a method and a provider that they feel comfortable with.

The voluntary nature of our proposed measure also significantly mitigates the potential negative impact on users’ rights to freedom of expression and association: users can continue to choose not to verify, and therefore retain the ability to
communicate and associate anonymously. Indeed users could have one verified account alongside an unverified account that they used when they felt the need for that privacy.

We should acknowledge that there would, however, still be some residual potential negative impact for users who chose not to verify. The combination of voluntary verification with labelling of verification status, and options for users to choose to filter interaction from non-verified accounts, would likely have some impact on the reach, and in some cases the credibility of users who chose anonymity. These features would disproportionately impact accounts that wished to misuse anonymity, for example to deceive or send unsolicited, hateful content. But there might be circumstances where a perfectly law-abiding user who has chosen not to verify for perfectly legitimate reasons finds it takes longer to build a following or reach an audience because of the filter, or where it takes longer for them to build credibility or overcome scepticism as a result of their non-verified status being visible.

If a verification system is not designed with accessibility in mind, there would be a risk that some of these negative impacts could fall disproportionately on already marginalised groups who might find it harder to complete a verification process. Whilst with a voluntary verification scheme the negative impact of being unable to access it is far less severe than with a mandatory scheme, platforms should still carefully consider equality and accessibility. Offering users a choice of methods of verification, including methods which will cater for otherwise marginalised groups, will mitigate these risks.

We would suggest this limited potential for a negative impact on some users who are unwilling or unable to verify must be balanced against the profound threat to fundamental rights and democratic values across the EU if the integrity of elections is not preserved. We would argue that the limited level of impact of an optional verification scheme appears proportionate to the aims pursued.

6. Conclusion

Fake accounts are a key enabler of Influence Operations. Whilst these operations evolve and innovate their tactics, techniques and procedures over time, a reliance on fake accounts to disseminate content, deceive users about its provenance, reliability and trustworthiness, and trick recommender algorithms into amplifying it is something of a constant.

An optional identity verification scheme, combined with clear and visible labels and optional filters, has the potential to disrupt Influence Operations’ use of the fake accounts, and significantly reduce their potency. It would do so in a way which
respects users’ fundamental rights, empowering them with greater information and greater choice, and avoiding any compulsory verification.

Such a measure would fit well with the measures which the Commission is already proposing, and in particular would strengthen the effectiveness of other suggested media literacy and contextual information measures. We would therefore encourage the Commission to consider adding this measure to its election integrity guidelines.