Emailed to: IHconsultation@ofcom.org.uk

27 February 2024

Consultation: Protecting People from Illegal Harms Online

Thank you for the opportunity to respond to Ofcom's consultation: Protecting People from Illegal Harms Online. In this response we have provided:

Section 1: Introduction to the Trust Alliance Group (Including the Communications Ombudsman and the Internet Commission)

Section 2: Our response to consultation questions

Section 1: Introduction to the Trust Alliance Group (Including the Communications Ombudsman and the Internet Commission)

Trust Alliance Group is a not-for-profit private limited company established in 2002 which runs a range of discrete national Alternative Dispute Resolution (ADR) schemes across different sectors, including the Ofgem-approved Energy Ombudsman and the Ofcom approved Communications Ombudsman.

Our purpose is to build, maintain and restore trust and confidence between consumers and businesses and we're developing diverse capabilities and expertise in a range of areas including digital alternative dispute resolution and case management technology.

The Internet Commission – a non-profit organisation which promotes ethical business practice to counter online harms whilst protecting privacy and freedom of expression and increase platform accountability – was acquired by the Trust Alliance Group in 2022.

The Internet Commission offers:

- Support to organisations who want to achieve high standards in online trust and safety
- Knowledge exchange where companies can discuss challenges and solutions related to tackling online harms
- A bank of good practices and reporting on the state-of-the art regarding governance and procedures of moderation of user-generated content online.

Section 2 – Our response to consultation questions

Question 6.1: Do you have any comments on Ofcom's assessment of the causes and impacts of online harms? Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.

We support Ofcom's assessment of the causes and impacts of online harms, which acknowledges the real-life impact lack of regulation, up until this point, has had on users online - particularly users who are vulnerable.

Question 6.2: Do you have any views about our interpretation of the links between risk factors and different kinds of illegal harm? Please provide evidence to support your answer.

The inclusion of user base risk factors effectively illustrates the potential value of a small user base to bad actors and we believe this approach should be reflected in other areas of the code – for instance, in relation to assessment of risk. However, elsewhere in the Code, it is stated that *"all else being equal, the more users a service has […] the greater the impact of any illegal content"*. This does not adequately capture the complexity of the situation and it would be beneficial to remain mindful of the serious harm that can be caused by smaller platforms. Other aspects such as the kind of users, their vulnerability, social context, etc. will impact the level of risk that platforms pose to individual users qualitatively, rather than just in terms of volume.

The inclusion of business model risk factors is something that we would argue should be more consistently represented throughout the codes. We would argue that business models would be worthy of inclusion within risk profiles, if only to highlight to companies the ways in which this factor may impact their service design and operation and to focus their attention towards this. Business models were a relevant factor in the Internet Commission's Accountability Reports, particularly regarding corporate governance and purpose.

The Internet Commission's Digital Accountability Reports identified that the business model of participant organisations had a significant impact on the implementation of trust and safety systems and tools. That is not to say that one or other model was better, but that it had a significant role in shaping what a service looked like, how its users were protected and how those users understood their role within the service.

It also meant that some services had greater incentive to minimise the number of child users. For example, if their business model relied on the paid subscriptions of adults who wanted to meet other adults (i.e. an online dating service), they were much more effective in leveraging age assurance methods. On the other hand, platforms relying on advertisers paying for the service, so that it could be delivered to users for free, would be far less incentivised to implement effective age assurance methods. This also intersected with their willingness to ban users; where the incentive to drive engagement, rather than maintain high community standards, meant volatile or incendiary users could be seen as an asset.

Ofcom should consider placing greater importance on the role of business models in informing the risk profiles of in-scope services and use it as a tool to inform its regulatory practices in evaluating submitted risk assessments.

Question 8.1: Do you agree with our proposals in relation to governance and accountability measures in the illegal content Codes of Practice? Please provide underlying arguments and evidence of efficacy or risks to support your view.

Yes, good governance by providers should lead to good risk management and mitigate risk and we support this being one of Ofcom's strategic priorities. This should result in a decrease in illegal content and harm.

Longer term, it may become apparent that user access to independent and expert Alternative Dispute Resolution (ADR) could be the missing piece of the puzzle, with regards to making online experiences safer. As with the Communications Ombudsman, such a service would be able to resolve individual disputes and also provide industry-level insights. Such insights could be a key to industrial and regulatory efforts to improve.

Trust Alliance Group
3300 Daresbury Park
Daresbury
Warrington WA4 4HS
Registered in England and Wales.
Company registration number: 04351294
VAT registration number: 798 3441 79

Although ADR is not currently a requirement of the Online Safety Act, we are pleased that a mechanism for the creation of an ADR duty is included at section 217.

Article 21 of the EU Digital Services Act (DSA) provides for the creation of out-of-court dispute settlement (ODS) bodies. As user-to-user services in the UK are the same as the EU, and UK citizens will have access to EU ODS bodies, it would make sense for Ofcom to promote consistency in the implementation of any future ADR requirement with the implementation of Article 21 of the DSA.

The DSA Observatory highlights that ODS does more than just resolve individual disputes. As part of the overall framework of the DSA, ODS bodies will be required to report on their activities, providing data and insights that will help identify systemic risks and harm. This will enable targeted regulatory interventions and mitigation measures to be put in place by platforms. Until ADR becomes a requirement of the OSA, equivalent data and insights will not be available to Ofcom.

Our experience operating ADR services in energy and communications markets leads us to believe that access to such a provision could offer:

- Independent redress for users to raise disputes
- A complete overview of emerging issues in digital markets
- The opportunity to spot issues of concern with individual platforms
- Clear and transparent categorisation of complaint types
- The capture of consumer experiences and detriment
- The ability to share information with platforms and regulators to drive improvements

TAG is developing our thinking and evidence base with regard to the provision of ADR in digital markets, and we will share this with Ofcom, if and when needed, to support a move to mandating ADR for UK digital markets.

Question 8.2: Do you agree with the types of services that we propose the governance and accountability measures should apply to?

We deem the list of types of services proposed to be sufficiently comprehensive. Our only doubt relates to video chat services (e.g. Omegle) which could technically fall out of scope of any types listed, due to the phrasing used.

Question 8.3: Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to requiring services to have measures to mitigate and manage illegal content risks audited by an independent third-party?

We welcome that independent third-party auditing is an option for providers in ensuring that measures taken to mitigate illegal harms are effective.

If a future change was being considered by Ofcom to make this a requirement rather than an option, TAG would be happy to share our experience of the Internet Commission's work as an independent third-party auditor and provide more information to show the efficacy, costs and risks of third-party auditing.

Trust Alliance Group
3300 Daresbury Park
Daresbury
Warrington WA4 4HS
Registered in England and Wales.
Company registration number: 04351294
VAT registration number: 798 3441 79

Question 9.2: Do you think the four-step risk assessment process and the Risk Profiles are useful models to help services navigate and comply with their wider obligations under the Act?

The approach is flexible, but it seems from other regulations and industry that this is preferable for both companies and regulators, given the variety of providers in-scope. We believe the recommendations of our 1.0 and 2.0 reports are suitably covered by this approach and we support the logic applied to the four-step process.

However, the four steps fail to account for the significant role of Safety by Design. Of course, it is not possible to implement Safety by Design measures for services or functionalities already in use, but service, product and functionality development are ongoing and should be prioritised within the risk assessment process.

Consideration of design, and its role in promoting safety online, should be part of the risk assessment process and should include those principles outlined by DSIT in its Principles of safer online platform design. By doing so, the risk assessments would be able to account for the ways in which services have, for example, empowered users to make safer choices and acknowledge the agency and influence of users on digital communities and experiences.

It should be the case that when companies are developing a new online service, product or functionality, they are able (or are required) to do so in line with a set of codes that ensures development is aligned with Safety by Design. This would have the added benefit of furnishing Ofcom with evidence about the ways in which Safety by Design is being implemented and allow Ofcom to maintain a view of innovation in the space – both in terms of new services but also new risk mitigation measures.

Regarding the reporting and review stage of the four-step process, it is not clear from the Code how the assessment of what is reported to Ofcom will take place, or how measurements will be made to generate comparable insights. Without the regulator being able to generate comparable insights and identify leading practices, it is unclear how companies or services will be able to accurately assess the efficacy of their systems and processes, or how to track improvements over time. This is particularly relevant to governance structures and accountability processes, which may be less overtly linked with quantitative measures of harm.

In the second of the Internet Commission's Accountability Reports, we identified 'Organisation, people and governance' as a key area of digital corporate responsibility. The report discussed how considerations of trust, safety and freedom are integrated into organisational culture and practice. In exploring this, we cited the following example of 'best practice', regarding high-level engagement with users:

"Twitch includes several high-profile content creators in its Safety Advisory Council. The Council was established to inform product and policy decisions and highlight the potential impacts on marginalised people. By including both online safety experts and the service's content creators, who deeply understand trust and safety challenges on the service, Twitch synthesises academic input with practical experience and better informs the development of safe environments online. Moreover, engaging content creators at this high level formalises the relationship between the organisation and its users and empowers the user community. Twitch advances digital responsibility by integrating community input into its organisation structure, addressing disconnections between the service's governance policies and the practical realities of its products and policies for the user community."

We used our maturity model[1] to situate an organisation's practice in relation to a scale running from "Elementary" to "Transformational", across which the practice may move over time (and was evidenced in participating across more than one assessment cycle). This is an example of a way in which improvements in qualitatively measured practices can be tracked and made valuable to both an overseeing body and other services, who can learn from best practice and be encouraged to develop their own approach.

Incentivisation to improve practices can be generated not only by highlighting enforcement measures and new guidance, but by emphasising positive outcomes for businesses and consumers.

We would question a regime where services can be compliant through implementation of recommended measures – even if that implementation has no impact or leads to no substantive mitigation of harm. Instead, we would suggest that promoting outcome-driven compliance would foster innovation in the development of alternative measures across the burgeoning safety tech industry.

Question 9.3: Are the Risk Profiles sufficiently clear and do you think the information provided on risk factors will help you understand the risks on your service?

The U2U Risk Profile is clear, although there are some issues we would like to see addressed.

When considering the first question of the Risk Profile we would ask why the file storage service type is described as concerning those services whose *"primary functionalities involve enabling users to store digital content and share access to that content through links".* This could allow services which only provide this functionality as a supplementary feature to argue that, since it is not their primary functionality, they do not need to identify such risks, thereby avoiding related duties.

There is also the issue of whether service type factors should include business model. In the Internet Commission's Accountability reports, it became clear that the business model of the participant organisation had a significant impact on the implementation of trust and safety systems and tools. That is not to say that one or other model was better, but that that it had a significant role in shaping what a service looked like, how its users were protected and how those users understood their role within the service.

It also meant that some services had greater incentive to minimise the number of child users. For example, if their business model relied on the paid subscriptions of adults who wanted to meet other adults (i.e. an online dating service), they were much more effective in leveraging age assurance methods. On the other hand, platforms relying on advertisers paying for the service so that it could be delivered to users for free would be far less incentivised to implement effective age assurance methods. This also intersected with their willingness to ban users; where the incentive to drive engagement, rather than maintain high community standards, meant volatile or incendiary users could be seen as an asset.

Ofcom should consider the role of business models in informing the risk profiles of in-scope services and use it as a tool to inform its regulatory practices in evaluating submitted risk assessments.

[1] Drawing on literature from the field of Corporate Social Responsibility (Głuszek, Ewa (2018) "Dimensions And Stages Of The CSR Maturity". Prace Naukowe Uniwersytetu Ekonomicznego We Wrocławiu, no. 520: 64-80. doi:10.15611/pn.2018.520.06.), we used an organisational maturity model (See Figure 2 on pg. 18 of our Accountability Report 2.0 available online) to evaluate an organisation's practices and its wider social impact. We asked, what does this practice tell us about the organisation's strategic approach and model for digital responsibility? We used a five-stage scale to evaluate the maturity of participating organisations by exploring and testing the congruence of observed practices and the organisation's digital responsibility goals.

Trust Alliance Group
3300 Daresbury Park
Daresbury
Warrington WA4 4HS
Registered in England and Wales.
Company registration number: 04351294
VAT registration number: 798 3441 79

The question *"Does my service allow child users to access some or all of the service?"* is too imprecise and could facilitate the under-reporting of children's presence on online services. It is well understood that children are present on services where they are not 'allowed' to be, that services for older children appeal to younger children and that many online services capitalise on this appeal – while simultaneously stressing they are making efforts to prevent younger children from accessing the service. The phrasing of the question – without any appeal for supporting evidence or follow-up questions about how the service limits inappropriate access – could be improved.

If this question remains unchanged, we suggest that a supplementary request for evidence is included. Alternatively, the question could be reframed to ask if the services *prevents* child users from accessing some or all of the service. The questions could also request information relating to the *means* by which this is achieved and what *portions of the service* are covered. Even without quantitative data about the rate of child users on a service, this additional information about child access could be very informative as to subsequent risk.

We would also suggest that another change be made in the risk profile to reflect the importance of user identity verification or assurance. The risk profile does not quite capture the risks presented by user base, despite user base being included as a risk factor in Volume 2 of the consultation documents. User base could be included as a section within the profile to reflect Ofcom's own thinking on the subject – incorporating the child users, age assurance, anonymity and the provision of identity verification services. Alternatively, it could be incorporated into Q4 of the risk profile ("Does my service have any of the following functionalities related to how users network with one another?"), under which identity verification could be included as a cross-cutting feature, impacting any type of user networking/interaction.

The reasoning behind this proposition is that the types of users significantly influence the nature, impact and prevalence of risks and harms on a platform, and this fact substantially intersects with the ways in which the size of a userbase influences risks and harms. For example:

- A suicide forum may only have a few hundred users, but those users are particularly vulnerable and inclined to share distinctly harmful content
- Snapchat has 500m users, but includes children of different age bands interacting
- Linkedin has 1bn users, but most are professional adults

Regarding children, one of the key challenges represented by platforms like this is that children of different age bands are mixing, which exacerbates the risk of harm as younger children look up to, want to emulate, and will 'obey' the older users of these platforms. This is in contrast to, say, PopJam where the userbase really was just comprised of one age band (see AADC age bands, for example.).

Ofcom should compel platforms to state what their target market is and provide information to show who is *actually* using the service. In this way, platforms will be able to make more informed assessments of the risk presented by the operation and delivery of their service. It may be the case that there is a gap between the users that the service is targeting and the users that are actually using the service. This would be valuable information for the service and for Ofcom to have and would allow for a more accurate understanding of the risks presented by the service. If the online service provider claims not to know who its users are or who is using the service, then this should indicate a higher risk level. It's important to note that there is a distinction between knowing who users are (e.g. IDed) compared to who, broadly, is using the service.

We stress that we are not advocating for uniform prescription of identity and age verification tools across the digital ecosystem, nor are we advocating for the kind of monitoring of personal characteristics or behavioural data that would empower a service to make deductions about its users. As an ADR provider, we are uniquely

Trust Alliance Group
3300 Daresbury Park
Daresbury
Warrington WA4 4HS
Registered in England and Wales.
Company registration number: 04351294
VAT registration number: 798 3441 79

situated to gain insight about service providers' customers for the purpose of providing information on systemic risks and harms. ADR is a vital tool in terms of effectively capturing and reporting data on user profiles and experiences and is an effective means of strengthening the regulatory frameworks of a number of industries by creating a regulatory feedback loop.

Additionally, 'audio' as a mode of communication seems to have been overlooked in Question 5 and Ofcom may deem it appropriate to include audio messaging as distinct from *"Posting or sending images or videos (either open or closed channels)"*. While there may be inadequate evidence to link it to the same harms as images or videos, it could still be included as another subtype and its potential for facilitating online harm should not be dismissed.

The role of AI generated content should also be factored into Question 5. While it may be that Ofcom considers it out of scope at this juncture, we have already seen the impact of generative AI in committing crimes online. One way to incorporate AI at this stage would be to ask whether there is a functionality enabling users to generate AI content within the service. It may also be worth considering whether the service has any integrations with other services which would facilitate generation of content in this way, or the dissemination of harmful content from another service. Through, for example, the EU's Digital Markets Act, interoperability will play a more significant role in the delivery of online services. Services will need to account for this when considering the risks that they and interoperating services may present.

Question 11.3: Do you agree with our definition of large services?

Ofcom's definition of large services is reasonable and its alignment with the EU DSA's threshold of 10% population is sensible.

However, we would question the placement of 'large services' as a defining category of online services alongside 'small services'. This binary representation of the digital landscape fails to capture and reflect the complexities of the market, including patterns of behaviour whereby bad actors deploy practices on smaller platforms and gradually develop them on larger ones until they are sufficiently sophisticated to evade the moderation practices of the largest services.

While reasonable in the abstract, the threshold for 'large users' is so high that it will not be met by companies like Roblox, which are among the largest platforms in the UK in use by children - one of the most vulnerable groups of Internet users. This means that very large online services, with a significant duty of care, fall short of being required to adhere to some of the more effective risk mitigation measures.

While those within the category of 'small services' will be subject to less stringent risk assessment requirements, the fact remains that 100,000 platforms are expected to be within scope of the Act, and it will be important for Ofcom to ensure it has sufficient resource to adequately evaluate the risk assessments submitted by online services.

It is possible that among the mass of smaller services, at least some may feel inclined not to identify risks present on their platforms, to avoid the designation of high or multi-risk – which would incur additional duties. It is not clear how this will be addressed, or whether greater attention will be paid to some services, more than others. For example, will Ofcom pay greater attention to risk assessments from services who are still very large, but just below the threshold? Or will all 'smaller' services be treated equally, outside of consideration of self-identified risk level? It seems that the use of 'large' services is a blunt instrument, and the details of its implementation could be clarified.

Trust Alliance Group
3300 Daresbury Park
Daresbury
Warrington WA4 4HS
Registered in England and Wales.
Company registration number: 04351294
VAT registration number: 798 3441 79

This binary approach risks overlooking the nuances of the relationships between multiple factors, including harm, size of services by userbase, size of a service by investment and capability and the way in which smaller services may facilitate more targeted and impactful harm (e.g. suicide fora). The assumption that "*all else being equal, the more users a service has, the more users can be affected by illegal content and the greater the impact of any illegal content"* is useful in terms of expediency but could lead to missing opportunities to capture and mitigate some of the most severe illegal harms perpetrated online. While a focus on larger services appears logical, it should be remembered that spread some of the most egregious illegal harms is spread by services who remain small and deliberately keep a low profile.

Please do not hesitate to contact us if you would like further information regarding our response. Our response is not confidential.

For more information regarding this response, please contact:

[✂]

Trust Alliance Group
3300 Daresbury Park
Daresbury
Warrington WA4 4HS
Registered in England and Wales.
Company registration number: 04351294
VAT registration number: 798 3441 79