

Your response

Please refer to the sub-questions or prompts in the [annex](#) to our call for evidence.

Question	Your response
<p>Question 1: Please provide a description introducing your organisation, service or interest in Online Safety.</p>	<p>Modulate provides the world’s only voice-native moderation service.</p> <p>ToxMod is designed first and foremost to ensure that (a) you learn about ALL the worst behavior happening on your platform, not merely that which players report to you; (b) you discover these harmful behaviors promptly enough that you can intervene (manually via moderation or automatically) during the session to prevent further escalation; (c) ToxMod preferentially flags misbehavior that genuinely harms others, rather than e.g. friendly trash talk; and (d) as your community evolves and new norms come into play, ToxMod will automatically learn about these changes and adapt its understanding of what constitutes a harm accordingly.</p> <p>ToxMod is able to achieve these goals thanks to its unique triaging system. When reviewing a voice chat session, ToxMod's initial analysis is geared to quickly differentiate between friendly and safe conversations and those which might warrant additional interest (this stage of analysis can also be done directly on the client device, before sending audio to ToxMod's servers, if preferred.) It then performs a deeper analysis on the flagged conversations, considering a massive number of factors including the content of the speech,</p>

	<p>its emotional nuance, how others in the chat are reacting, speech behaviors like crying or laughing, signals you provide such as player reports or in-game behavior, each player's history with different types of misbehavior, and potential demographic differences between players (e.g. a child speaking to an adult.) The end result of this analysis is that ToxMod shares with you any instance it determines to be problematic alongside a category (Sexual Harassment, Adult Language, Violent Speech, etc) and a score indicating the severity of the offense. You can configure ToxMod to present these takeaways to your moderators alongside additional context (ensuring your moderation team makes the final call) or to automatically send the data regarding these flagged offenses to your player management system to enable automated bans on the most severe offenses.</p> <p><i>Is this response confidential? – N</i></p>
<p>Question 2: Can you provide any evidence relating to the presence or quantity of illegal content on user-to-user and search services?</p> <p>IMPORTANT: Under this question, we are not seeking links to or copies/screenshots of content that is illegal to hold, such as child sexual abuse. Deliberately viewing such images may be a criminal offence and will be reported to the police.</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>
<p>Question 3: How do you currently assess the risk of harm to individuals in the UK from illegal content presented by your service?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>
<p>Question 4: What are your governance, accountability and decision-making structures for user and platform safety?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>

<p>Question 5: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements?</p>	<p>Establish shared vocabulary and standards that are consistent across all platforms, similar to nutrition labels on food; these are not always perfect (“No sodium” still can contain trace amounts of sodium) but make it extremely effective for end users to quickly identify what they care about and how they are meant to use the product.</p> <p><i>Is this response confidential? – N</i></p>
<p>Question 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?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>
<p>Question 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?</p>	<p>Notify users clearly when they are being banned and explain why their behavior was inappropriate; this feedback has proven extremely effective in games such as Apex Legends in reducing recidivism in games, and also builds trust with end users who understand that action is actually being taken.</p> <p><i>Is this response confidential? – N</i></p>
<p>Question 8: If your service has <i>reporting or flagging</i> mechanisms in place for illegal content, or users who post illegal content, how are these processes designed and maintained?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>
<p>Question 9: If your service has a <i>complaints</i> mechanism in place, how are these processes designed and maintained?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>

<p>Question 10: What action does your service take in response to reports or complaints?</p>	<p><i>Is this response confidential? – Y/N (delete as appropriate)</i></p>
<p>Question 11: Could improvements be made to content moderation to deliver greater protection for users, without unduly restricting user activity? If so, what?</p>	<p>Yes, absolutely. Basically no voice content is moderated today, yet it's essential for reliably building strong social connections. Moderating voice chat is a bare minimum requirement that every platform should be focusing on, and requires no alterations to player behaviour.</p> <p>It should further be noted that "restricting player behaviour" is typically not actually tied to having content moderation, it's tied to <i>what the Code of Conduct is that is actually being moderated on</i>.</p> <p>It's possible for platforms to have an overly restrictive Code of Conduct (though one might argue that's their right)...but most platforms today have pretty unrestrictive Codes of Conduct, yet even then, don't actually hold users to the Code of Conduct because they don't have the moderation tools to do so.</p> <p><i>Is this response confidential? – N</i></p>
<p>Question 12: What automated moderation systems do you have in place around illegal content?</p>	<p><i>Is this response confidential? – Y/N (delete as appropriate)</i></p>
<p>Question 13: How do you use human moderators to identify and assess illegal content?</p>	<p><i>Is this response confidential? – Y/N (delete as appropriate)</i></p>
<p>Question 14: How are sanctions or restrictions around access (including to both the service and to particular content) applied by providers of online services?</p>	<p><i>Is this response confidential? – Y/N (delete as appropriate)</i></p>

Question 15: In what instances is illegal content removed from your service?	<i>Is this response confidential? – Y/N (delete as appropriate)</i>
Question 16: Do you use other tools to reduce the visibility and impact of illegal content?	<i>Is this response confidential? – Y/N (delete as appropriate)</i>
Question 17: What other sanctions or disincentives do you employ against users who post illegal content?	<i>Is this response confidential? – Y/N (delete as appropriate)</i>
Question 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?	<ol style="list-style-type: none">1. Proactive moderation (e.g. not waiting for user reports, but taking action directly based on manual or automated oversight.) This already exists in text but is heavily neglected in other areas such as audio and video. The Anti-Defamation League found that only 8% of gamers ever report toxicity they encounter; the rate is almost certainly lower for illegal content like child predation, as the victim in such cases (the child) may not even know that what is happening ought to be reported. As such, proactivity is an absolute must.2. As mentioned earlier, Apex Legends found a massive decrease in recidivism through simply explaining to penalized individuals why their behaviour was problematic. Too few platforms actually explain their bans; this sort of simple change could be hugely powerful. <p><i>Is this response confidential? – N</i></p>

<p>Question 19: To what extent does your service encompass functionalities or features designed to mitigate the risk or impact of harm from illegal content?</p>	<p>ToxMod is specifically designed to detect and identify harmful behavior, including illegal content, in voice chat or voice streams.</p> <p><i>Is this response confidential? – N</i></p>
<p>Question 20: How do you support the safety and wellbeing of your users as regards illegal content?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>
<p>Question 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?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>
<p>Question 22: What age assurance and age verification technologies are available to platforms, and what is the impact and cost of using them?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>
<p>Question 23: Can you identify factors which might indicate that a service is likely to attract child users?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>
<p>Question 24: Does your service use any age assurance or age verification tools or related technologies to verify or estimate the age of users?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>
<p>Question 25: If it is not possible for children to access your service, or a part of it, how do you ensure this?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>
<p>Question 26: What information do you have about the age of your users?</p>	<p><i>Is this response confidential? – Y/ N (delete as appropriate)</i></p>

Question 27: For purposes of transparency, what type of information is useful/not useful? Why?	<i>Is this response confidential? – Y/ N (delete as appropriate)</i>
Question 28: Other than those in this document, are you aware of other measures available for mitigating risk and harm from illegal content?	<i>Is this response confidential? – Y/ N (delete as appropriate)</i>

Please complete this form in full and return to OS-CFE@ofcom.org.uk