Question	Your response
Question 1: Please provide a description introducing your organisation, service or interest in Online Safety.	Is this response confidential? – N Chayn is a nonprofit that creates digital, multilingual resources to support the healing of survivors of gender-based violence. Our focus is empowering women and other marginalised genders who have experienced domestic, sexual, or tech-based abuse. In 2013, Chayn was one of the first organisations of its kind to bridge the gap between gender and tech to create openly-sourced and licensed resources tackling gender-based violence. There was a serious gap online in intersectional resources on legal rights, tech abuse, and wellbeing. Women were turning to their browsers for answers and not finding them, and were often misinformed about their legal rights to deter them from seeking justice. The emerging and growing threat of tech abuse meant that seeking support was often unsafe. But the internet is for everyone. Designing accessible, trauma-informed, intersectional online resources that are safe for survivors to access is what guides our work. And we know that digital spaces can be the site of not only harm, but also healing. As a survivor-led organisation, every decision we make – and every resource we create – has lived experience at its core. Last year, Chayn partnered with End Cyber Abuse to go on a journey to understand the nature of technology facilitated gender-based violence and how we can address it. The result is Orbits: a guide on how we can design interventions to tech abuse that are intersectional, survivor-centred, and trauma-informed. Co-created with thinkers, practitioners, and survivors from around the world, the guide focuses on three areas that are vital for effectively tackling tech abuse: technology, research, and policy.
Question 2: Can you provide any evidence relating to the presence or quantity of illegal content on user-to-user and search services? 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	Is this response confidential? – Y

may be a criminal offence and will be reported to the police.	
Question 3: How do you currently assess the risk of harm to individuals in the UK from illegal content presented by your service?	Is this response confidential? – Y / N
Question 4: What are your governance, accountability and decision-making structures for user and platform safety?	Is this response confidential? – N Our platforms and services go through multiple iterations of consequence scanning for potential safety concerns, as well as user testing with survivors of abuse. Our Head of Product and Experience, alongside our project teams and User Researcher, are responsible for overseeing these product builds and continuously assessing our products for further security needs. Our Tech Engineers secure the digital frontier of our online products, ensuring that user data is secure and confidential at all times. Most of all, our products are created under a trauma-informed design principles approach that enables us to reduce harm from the beginning. We have documented these processes for the benefit of others. Safety is not just a necessary product feature; it's a fundamental aspect of providing trauma-informed care. We know that survivors will have different online safety needs and concerns as they access our resources - some may still be monitored by an abuser, for example. As such, we employ a variety of safety features across our platforms; for example, we have an 'exit button' on all our platforms, for users who need to conceal their web activity from people in their physical environment. To users of our online courses, we advise them on how to delete our videos from their YouTube watching history. Moreover, we have a guide on online safety (https:// www.chayn.co/safety) in 7 languages, and continuously signpost different support services throughout our resources. Because if a user is unable to access a service due to safety concerns, that service has failed in its mandate of accessibility.
Question 5: What can providers of online services do to enhance the clarity and	Is this response confidential? – N

accessibility of terms of service and public policy statements?	**Use of accessible language.** We've all read inscrutable terms and conditions that employ obscure legal terminology. The less intelligible the policy, the more services wilfully obscure their commitment to ending online abuse. When Chayn performs user research interviews, in addition to providing a consent form, we also provide a video explaining the consent form to survivors. This makes it more understandable and accessible to those who find it hard to read through lots of text. **Translation into multiple languages,** in particular those spoken by migrant populations in the UK, e.g., Arabic, Hindi, Somali, and Pashto. Work that is not diverse by design will be unequal in outcome. Tech platforms must be designed for use by all, which includes terms of service. Increasing the language accessibility of policy and terms of service will increase the equitable design of the platform and ensure all users have an equal chance at participation. **Comprehensive user testing**. Services should work with users from diverse backgrounds to test the comprehensibility of any written material they produce regarding use of their platform. **Exemplify harmful behaviours while highlighting policies on zero tolerance of abuse.** Without reproducing illegal or harmful content, specificity to the kinds of behaviour that are not tolerated on the platform will serve to clarify terms of service and clearly warn users against harmful behaviours.
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?	Is this response confidential? – Y / N
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?	Is this response confidential? – N Release reporting figures publicly, and provide up-front, realistic, data-led expectations about reporting procedures, e.g., average response time, and the frequency of follow-up

	messaging. Offer full transparency of the reporting procedure, including communicating to survivors which department deals with the report, and informing them that there is a dedicated and specialist team who handles reports. Permit third-party reporting. Permit users to report other users' offline behaviour. Provide comprehensive reporting mechanisms that let survivors report even if the perpetrator deactivates/disconnects their account. Create flexible mechanisms that enable people to describe their own experience and share the remedial measures they wish for, rather than forcing reports into rigid, predetermined categories. Provide information on reporting mechanisms in multiple formats, e.g., videos as well as written text. Implement consent at various stages of the reporting processes. This means actively asking survivors for their consent in sharing information with other agencies and individuals within the organisation, and being clear with survivors about how and why their information is being shared, and why their information is being shared. Allow users to identify multiple offences in one complaint. An intersectional reporting procedure will account for how some users may be simultaneously subjected to multiple forms of hate and/or oppression, e.g., racism and transphobia, at the same time. As such, users should be able to report user actions as breaching multiple terms of service. Allow users to make a report in their local language. This would also mean hiring content moderators who have proficiency in a variety of languages, and are able to respond to user complaints in a culturally-sensitive manner, taking context into account. This also involves

	making policies and reporting mechanisms available in different languages and dialects. Design a reporting procedure which does not necessitate re-exposure to the harmful content. Having to view the reported content multiple times throughout the reporting procedure is re-traumatising for users, and acts as a disincentive to report in the first place. Facilitating a less traumatising reporting experience will result in more reports and hence greater user safety. Automatically disable cookies and tracking when survivors report abuse on platforms. Offer reporting processes with accessibility considerations embedded, including an option for low-bandwidth or offline reporting. Enable the ability to use voice in the reporting process. Users with limited literacy or limited ability in the platform's language should be able to submit user reports via voice. Lengthen the timeframe during which abuse can be reported. Users may not feel safe to report right when the abuse initially occurs, or may be experiencing trauma reactions from the abuse and feel unable to report in its immediate aftermath. Ask users for safe contact details as part of the reporting process, as these may differ from the details under which they registered their account, for example if their online activity is being monitored by an abuser. Implement quick access bars for reporting abuse alongside a dedicated Safety Centre within each platform, that allows users to track the progress of the report, even after the report has been marked as complete. Give upfront information on the circumstances under which a user report will be handed over to law enforcement, and what information will be shared in such a case. Many users may have safety concerns related to their information being given to police, for example those from a migrant background whose status in a country

	is not secure, and clear information on any limits of confidentiality must be given upfront to encourage reporting. Offer interactive online modules on the terms of service, including unacceptable and inappropriate behaviour, so users are empowered and informed to identify when they are experiencing or witnessing online harm. Ensure confidentiality of survivors' personal details while reporting instances of abuse on tech platforms or with law enforcement agencies. Signpost additional forms of support throughout the reporting process. Re-traumatisation is frequently the reason survivors choose not to report abuse. Providing survivors with a digital file of evidence that can support civil and criminal cases, if they want to pursue those routes.
Question 8: If your service has reporting or flagging mechanisms in place for illegal content, or users who post illegal content, how are these processes designed and maintained?	Is this response confidential? – Y / N
Question 9: If your service has a complaints mechanism in place, how are these processes designed and maintained?	Is this response confidential? – Y / N
Question 10: What action does your service take in response to reports or complaints?	Is this response confidential? – Y / N
Question 11: Could improvements be made to content moderation to deliver greater protection for users, without unduly restricting user activity? If so, what?	Is this response confidential? – N Establish reporting mechanisms that don't involve further distributing harmful content, and do not involve re-exposure of the harmful content to the user who made the report. Allow users to customise settings for who can contact them via direct message, who can send them media, and how their material can be shared and downloaded. Employ digital fingerprinting to assist with removing offending materials from all platforms and flagging accounts that shared the offending materials.

	Train content moderators in trauma-informed care and cultural context so they are able to accurately identify harmful content and communicate with users sensitively. This should also involve hiring diverse content moderators with an understanding of a variety of languages and cultural contexts. Allow users to appoint a trusted user who can block, report, and moderate content and/or other users on the user's behalf. Employ gender-inclusive terms of service, for example having gender-free regulations on permitted images (see the Gender Inclusive Content Moderation report here).
Question 12: What automated moderation systems do you have in place around illegal content?	Is this response confidential? – Y / N
Question 13: How do you use human moderators to identify and assess illegal content?	Is this response confidential? – Y / N
Question 14: How are sanctions or restrictions around access (including to both the service and to particular content) applied by providers of online services?	Is this response confidential? – N **Age verification processes.** A variety of age verification processes exist, for example presentation of government-issued identity documents, use of biometric data, presentation of bank or mobile phone records, and more. **Identity verification processes.** These can be achieved for example by requiring users to present official identity documents at sign-up. Relatedly, some service providers employ name standards (e.g., 'real name' policies). **Two-factor authentication.** Some online service providers may require authentication with another device, for example a mobile phone. Additional requirements can be employed to verify user characteristics, for example 2-factor authentication with a mobile number that has an area code in the country where the service is registered and/or provided. **Restricting feature access,** for example some content only being visible to users who are signed in and whose age has been verified.
Question 15: In what instances is illegal content removed from your service?	Is this response confidential? – Y / N
Question 16: Do you use other tools to reduce the visibility and impact of illegal content?	Is this response confidential? – Y / N

Question 17: What other sanctions or disincentives do you employ against users who post illegal content?	Is this response confidential? – Y / N
Question 18: Are there any functionalities or design features which evidence suggests can effectively prevent harm, and could or should be deployed widely by industry?	Is this response confidential? – N Digital fingerprinting can assist with removing offending materials from all platforms and flagging accounts that shared the offending materials. Sharing last known logins, so survivors of abuse can spot if an abuser or stalker has managed to get control of their devices or accounts. Flagging and/or blur offensive content, and creating digital fingerprints to block uploading of flagged content. Customisable settings that allow users to: control how their images and other media can be downloaded and shared, who can be in touch with them and how, and whether their user profile is searchable and/or suggested to other users (e.g., ability to opt out of 'people you may know' functionalities). Using privacy-enhancing technologies (PET) such as encryption and data masking.
Question 19: To what extent does your service encompass functionalities or features designed to mitigate the risk or impact of harm from illegal content?	Is this response confidential? – Y / N
Question 20: How do you support the safety and wellbeing of your users as regards illegal content?	Is this response confidential? – N We have an Online Safety Guide with direct guidance on protecting ourselves and our data online, including protection against image-based abuse and other illegal distribution of content; it has been accessed by over 83,000 people since its launch in 2016. We are soon launching a course on image-based abuse as part of our Bloom project. This course will support survivors of image-based abuse in their healing, as well as explain their options for reporting abuse and protecting themselves online. The internet is for everyone; while women and people from marginalised backgrounds are more likely to experience online harm, we envision a future where all are not only protected but empowered on the

	internet as leaders, creators, and community builders. Protection against illegal content is only the first step; redressing harm also requires us to re-create the online spaces we inhabit to promote equity, safety, and inclusion.
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?	Is this response confidential? – Y / N
Question 22: What age assurance and age verification technologies are available to platforms, and what is the impact and cost of using them?	Is this response confidential? – Y / N
Question 23: Can you identify factors which might indicate that a service is likely to attract child users?	Is this response confidential? – Y / N
Question 24: Does your service use any age assurance or age verification tools or related technologies to verify or estimate the age of users?	Is this response confidential? – Y / N
Question 25: If it is not possible for children to access your service, or a part of it, how do you ensure this?	Is this response confidential? – Y / N
Question 26: What information do you have about the age of your users?	Is this response confidential? – Y / N
Question 27: For purposes of transparency, what type of information is useful/not useful? Why?	Is this response confidential? – N Transparency on reporting mechanisms: how long they take, all of the agents who will have access to the report, what the steps and contingencies of the reporting process are, and what happens to a report after it has been marked complete. Information about data usage: the conditions under which agents working on behalf of the platform have access to individual user data, and conditions under which individual user data will be shared with third-party agencies, for example law enforcement. Comprehensive, comprehensible information on all the conditions in which user data is and is not encrypted. In our view, any and all information pertaining to the use of individual user data, who can view it and when, and the conditions under which it

	is deaggregated is useful. The core issue at present is not that extraneous information is presented to users, but the fact that this information is often presented in inaccessible, formal terminology, and only once during the user signup process.
Question 28: Other than those in this document, are you aware of other measures available for mitigating risk and harm from illegal content?	Is this response confidential? – N A hugely under acknowledged issue in the tech industry is the mitigation of harm for content moderators. Content moderation is often outsourced to poorly paid and supported 'ghost workers', usually based in the Global South. Reviewing abusive content can be traumatising, yet these workers are rarely given sufficient training or psychological support. This practice extends, rather than mitigates, harm, and perpetuates colonial inequities and violence. Fair compensation and adequate access to trauma counselling and other forms of occupational support should be mandatory for content moderators exposed to illegal content on platforms.